GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 19:31:38 ; Search time 4930.81 Seconds
(without alignments)
17544.846 Million cell updates/sec

Title: US-09-301-380-1
Perfect score: 4134
Sequence: 1 cttcaaagttccccgcatga.....tgggtagagagtatattttc 4134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 799276

Minimum DB seq length: 15
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
c 1	28	0.7	29	6	AX099610	AX099610 Sequence
2	27.2	0.7	87	14	ROTVP7AK	M37350 Homo sapien
3	27.2	0.7	87	14	ROTVP7AL	M37351 Homo sapien
4	25.8	0.6	89	4	AF213231	AF213231 Bison bis
5	25.6	0.6	87	14	ROTVP7AJ	M37349 Homo sapien
6	25.6	0.6	87	14	ROTVP7AM	M37352 Homo sapien
7	25.6	0.6	87	14	ROTVP7AN	M37353 Homo sapien
8	25.4	0.6	93	4	AF213232	AF213232 Bison bis
c 9	25.4	0.6	94	5	AF272967	AF272967 Mastacemb
10	25	0.6	70	6	I71468	I71468 Sequence 9
c 11	24.8	0.6	97	5	AF174509	AF174509 Torrentop
c 12	24.8	0.6	97	5	AF174510	AF174510 Torrentop
c 13	24.6	0.6	90	12	SYNDEL951	M10780 X.laevis DN
14	24.4	0.6	70	8	AF326227	AF326227 Arabidops
15	24.4	0.6	86	1	S45981	S45981 Selenoprote
c 16	24.4	0.6	92	5	XELMYHCA	M27234 Xenopus lae
17	24.2	0.6	84	10	S79629	S79629 HSF2- alpha
18	24.2	0.6	95	3	AGXH454	Z72086 A.gambiae s
19	24.2	0.6	95	6	AX135308	AX135308 Sequence
c 20	24.2	0.6	100	11	HUMUT931A	L31299 Human STS U
21	24	0.6	87	14	ROTVP7AI	M37348 Homo sapien
22	24	0.6	87	14	ROTVP7AO	M37354 Homo sapien
23	24	0.6	92	6	AX192611	AX192611 Sequence
c 24	23.8	0.6	90	12	SYNDEL952	M10781 X.laevis DN
c 25	23.8	0.6	94	5	AF272954	AF272954 Monopteru
c 26	23.8	0.6	94	5	AF272968	AF272968 Mastacemb
27	23.8	0.6	98	8	OSA307923	AJ307923 Oryza sat
c 28	23.6	0.6	67	6	A60774	A60774 Sequence 83
c 29	23.6	0.6	90	3	DMADHAESA	X66233 D.melanogas
c 30	23.6	0.6	95	4	AF247752	AF247752 Gulo gulo
c 31	23.6	0.6	95	6	AX283217	AX283217 Sequence
32	23.6	0.6	97	9	HUM13COL15	M68981 Human alpha
33	23.4	0.6	66	6	AR068708	AR068708 Sequence
34	23.4	0.6	66	6	AR156638	AR156638 Sequence
35	23.4	0.6	66	6	E15527	E15527 Primer. 7/1
36	23.4	0.6	81	6	A21367	A21367 Repeat nucl
37	23.4	0.6	97	4	AF213233	AF213233 Bison bis
c 38	23.2	0.6	72	6	AX036609	AX036609 Sequence
c 39	23.2	0.6	96	6	AX103760	AX103760 Sequence
40	23.2	0.6	96	8	CNS01AFO	AL112876 Botrytis
c 41	23.2	0.6	99	6	A13167	A13167 Artificial
42	23	0.6	59	10	MMIIPKR08	U09920 Mus musculu
43	23	0.6	74	9	HUMIGDJG	L35136 Human IG pr
44	23	0.6	79	11	AF084216	AF084216 Mus muscu
c 45	23	0.6	90	12	SYNDEL41	M10778 X.laevis DN

ALIGNMENTS

RESULT 1
AX099610/c
LOCUS AX099610 29 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 250 from Patent WO0119988.
ACCESSION AX099610
VERSION AX099610.1 GI:13538665
KEYWORDS .
SOURCE synthetic construct.
ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 29)
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Collins-Racie,L.A., Evans,C., Merberg,D., Treacy,M., Bowman,M.R., Spaulding,V. and Agostino,M.J.
TITLE Secreted proteins and polynucleotides encoding them
JOURNAL Patent: WO 0119988-A 250 22-MAR-2001;
Genetics Institute, Inc. (US)
FEATURES Location/Qualifiers
source 1. .29
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="oligonucleotide"
misc_feature 2
/note="biotinylated phosphoaramidite residue"

```
BASE COUNT       7 a      9 c      4 g      8 t      1 others
ORIGIN


  Query Match             0.7%;  Score 28;  DB 6;  Length 29;
  Best Local Similarity   96.6%;  Pred. No. 2.9e+04;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      75 aaggaaattcagtgtgtgagtctcagcag 103
           ||||||||||||||||||||||||||||| |
Db      29 AAGGAAATTCAGTGTGTGAGTCTCAGCNG 1


RESULT   2
ROTVP7AK
LOCUS       ROTVP7AK                  87 bp ss-RNA     linear   VRL 23-SEP-1994
DEFINITION  Homo sapiens rotavirus 2 (strain 2743) outer capsid protein (VP7)
            RNA, partial cds.
ACCESSION   M37350
VERSION     M37350.1  GI:548268
KEYWORDS    VP7 gene; outer capsid protein.
SOURCE      Human rotavirus 2 (strain 2743) RNA.
  ORGANISM  Human rotavirus 2
            Viruses; dsRNA viruses; Reoviridae; Rotavirus.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Green,K.Y., Sears,J.F., Taniguchi,K., Midthun,K., Hoshino,Y.,
            Gorziglia,M., Nishikawa,K., Urasawa,S., Kapikian,A.Z., Chanock,R.M.
            and Flores,J.
  TITLE     Prediction of human rotavirus serotype by nucleotide sequence
            analysis of the VP7 protein gene
  JOURNAL   J. Virol. 62, 1819-1823 (1988)
  MEDLINE   88188272
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Human rotavirus 2"
                     /strain="2743"
                     /db_xref="taxon:36429"
     gene            1. .87
                     /gene="VP7"
     CDS             <1. .>87
                     /gene="VP7"
                     /codon_start=1
                     /product="outer capsid protein"
                     /protein_id="AAA47369.1"
                     /db_xref="GI:548269"
                     /translation="AEAKNEISDDEWENTKTTDVSTFEIVASS"
BASE COUNT       32 a     11 c     22 g     22 t
ORIGIN


  Query Match             0.7%;  Score 27.2;  DB 14;  Length 87;
  Best Local Similarity   61.1%;  Pred. No. 5.2e+04;
  Matches   44;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     818 ctgtgaaggtgatttcagtggatgaattgaatgacactatagctgctaatttgagtgaca 877
           ||   || | ||||||||  ||||||| | |  | |||| | || |  |  |||||
Db       8 CTAAAAATGAGATTTCAGATGATGAATGGGAAAATACTAAAACTACGGACGTGAGTACAT 67

Qy     878 ctgagttttatg 889
           |||| |   ||
Db      68 TTGAGATAGTTG 79


RESULT   3
ROTVP7AL
LOCUS       ROTVP7AL                  87 bp ss-RNA     linear   VRL 23-SEP-1994
DEFINITION  Homo sapiens rotavirus 2 (strain 2899) outer capsid protein (VP7)
            RNA, partial cds.
ACCESSION   M37351
VERSION     M37351.1  GI:548270
KEYWORDS    VP7 gene; outer capsid protein.
SOURCE      Human rotavirus 2 (strain 2899) RNA.
  ORGANISM  Human rotavirus 2
            Viruses; dsRNA viruses; Reoviridae; Rotavirus.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Green,K.Y., Sears,J.F., Taniguchi,K., Midthun,K., Hoshino,Y.,
            Gorziglia,M., Nishikawa,K., Urasawa,S., Kapikian,A.Z., Chanock,R.M.
            and Flores,J.
  TITLE     Prediction of human rotavirus serotype by nucleotide sequence
            analysis of the VP7 protein gene
  JOURNAL   J. Virol. 62, 1819-1823 (1988)
  MEDLINE   88188272
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Human rotavirus 2"
                     /strain="2899"
                     /db_xref="taxon:36429"
     gene            1. .87
                     /gene="VP7"
     CDS             <1. .>87
                     /gene="VP7"
                     /codon_start=1
                     /product="outer capsid protein"
                     /protein_id="AAA47370.1"
                     /db_xref="GI:548271"
                     /translation="AEAKNEISDDEWENTKTTDVSTFEIVASS"
BASE COUNT       32 a     11 c     22 g     22 t
ORIGIN


  Query Match             0.7%;  Score 27.2;  DB 14;  Length 87;
  Best Local Similarity   61.1%;  Pred. No. 5.2e+04;
  Matches   44;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     818 ctgtgaaggtgatttcagtggatgaattgaatgacactatagctgctaatttgagtgaca 877
           ||   || | ||||||||  ||||||| | |  | |||| | || |  |  |||||
Db       8 CTAAAAATGAGATTTCAGATGATGAATGGGAAAATACTAAAACTACGGACGTGAGTACAT 67

Qy     878 ctgagttttatg 889
           |||| |   ||
Db      68 TTGAGATAGTTG 79


RESULT   4
AF213231
LOCUS       AF213231                  89 bp    DNA     linear   MAM 09-FEB-2001
DEFINITION  Bison bison clone 15.1.1 microsatellite BMS410-89.
ACCESSION   AF213231
VERSION     AF213231.1  GI:11837834
KEYWORDS    .
SOURCE      American bison.
  ORGANISM  Bison bison
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bison.
REFERENCE   1  (bases 1 to 89)
  AUTHORS   Schnabel,R.D., Ward,T.J. and Derr,J.N.
  TITLE     Validation of 15 microsatellites for parentage testing in North
            American bison, Bison bison and domestic cattle
  JOURNAL   Anim. Genet. 31 (6), 360-366 (2000)
  MEDLINE   21070477
REFERENCE   2  (bases 1 to 89)
  AUTHORS   Schnabel,R.D., Ward,T.J. and Derr,J.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1999) Veterinary Pathobiology, Texas A&M
            University, Bldg. 1197 Rm.226, College Station, TX 77843-4467, USA
FEATURES             Location/Qualifiers
     source          1. .89
                     /organism="Bison bison"
                     /db_xref="taxon:9901"
                     /clone="15.1.1"
     primer_bind     1. .22
     repeat_region   1. .89
                     /note="microsatellite BMS410-89"
                     /rpt_type=tandem
     primer_bind     70. .89
```

```
BASE COUNT       29 a     33 c      8 g     19 t
ORIGIN


  Query Match             0.6%;  Score 25.8;  DB 4;  Length 89;
  Best Local Similarity   56.5%;  Pred. No. 1.3e+05;
  Matches   48;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy   1450 tttgtaaatgtgctggctgagccaccacgaatcctcacacctgcaaacacactctaccag 1509
          ||   | || | |   ||     | |||| |  | |||||   || |||||| |   ||
Db      1 TTGCCACATTTACCTTCTTTCATAGCACGCACACACACACACACACACACACACACAC 60

Qy   1510 gtcattgcaaacaggcctgctttac 1534
          ||  ||  |||||   | ||||| |
Db     61 GTGGTTATAAACACCACAGCTTTTC 85


RESULT   5
ROTVP7AJ
LOCUS       ROTVP7AJ                  87 bp ss-RNA     linear   VRL 23-SEP-1994
DEFINITION  Homo sapiens rotavirus 2 (strain HN126) outer capsid protein (VP7)
            RNA, partial cds.
ACCESSION   M37349
VERSION     M37349.1  GI:548266
KEYWORDS    VP7 gene; outer capsid protein.
SOURCE      Human rotavirus 2 (strain HN126) RNA.
  ORGANISM  Human rotavirus 2
            Viruses; dsRNA viruses; Reoviridae; Rotavirus.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Green,K.Y., Sears,J.F., Taniguchi,K., Midthun,K., Hoshino,Y.,
            Gorziglia,M., Nishikawa,K., Urasawa,S., Kapikian,A.Z., Chanock,R.M.
            and Flores,J.
  TITLE     Prediction of human rotavirus serotype by nucleotide sequence
            analysis of the VP7 protein gene
  JOURNAL   J. Virol. 62, 1819-1823 (1988)
  MEDLINE   88188272
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Human rotavirus 2"
                     /strain="HN126"
                     /db_xref="taxon:36429"
     gene            1. .87
                     /gene="VP7"
     CDS             <1. .>87
                     /gene="VP7"
                     /codon_start=1
                     /product="outer capsid protein"
                     /protein_id="AAA47368.1"
                     /db_xref="GI:548267"
                     /translation="TEAKNEISDDEWENTKTTDVNTFEIVASS"
BASE COUNT       34 a     11 c     19 g     23 t
ORIGIN


  Query Match             0.6%;  Score 25.6;  DB 14;  Length 87;
  Best Local Similarity   59.7%;  Pred. No. 1.4e+05;
  Matches   43;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    818 ctgtgaaggtgatttcagtggatgaattgaatgacactatagctgctaatttgagtgaca 877
          ||   || | ||||||||  |||||||| | |  | |||| | || |  || | |
Db      8 CTAAAAATGAGATTTCAGATGATGAATGGGAAAATACTAAAACTACGGATGTAAACACAT 67

Qy    878 ctgagttttatg 889
          ||||  ||  ||
Db     68 TTGAGATTGTTG 79


RESULT   6
ROTVP7AM
LOCUS       ROTVP7AM                  87 bp ss-RNA     linear   VRL 23-SEP-1994
DEFINITION  Homo sapiens rotavirus 2 (strain C) outer capsid protein (VP7) RNA,
            partial cds.
ACCESSION   M37352
VERSION     M37352.1  GI:548272
KEYWORDS    VP7 gene; outer capsid protein.
SOURCE      Human rotavirus 2 (strain C) RNA.
  ORGANISM  Human rotavirus 2
            Viruses; dsRNA viruses; Reoviridae; Rotavirus.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Green,K.Y., Sears,J.F., Taniguchi,K., Midthun,K., Hoshino,Y.,
            Gorziglia,M., Nishikawa,K., Urasawa,S., Kapikian,A.Z., Chanock,R.M.
            and Flores,J.
  TITLE     Prediction of human rotavirus serotype by nucleotide sequence
            analysis of the VP7 protein gene
  JOURNAL   J. Virol. 62, 1819-1823 (1988)
  MEDLINE   88188272
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Human rotavirus 2"
                     /strain="C"
                     /db_xref="taxon:36429"
     gene            1. .87
                     /gene="VP7"
     CDS             <1. .>87
                     /gene="VP7"
                     /codon_start=1
                     /product="outer capsid protein"
                     /protein_id="AAA47371.1"
                     /db_xref="GI:548273"
                     /translation="AEAKNEISDDEWENTKTTDVNTFEIVASS"
BASE COUNT       33 a     11 c     20 g     23 t
ORIGIN


  Query Match             0.6%;  Score 25.6;  DB 14;  Length 87;
  Best Local Similarity   59.7%;  Pred. No. 1.4e+05;
  Matches   43;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    818 ctgtgaaggtgatttcagtggatgaattgaatgacactatagctgctaatttgagtgaca 877
          ||   || | ||||||||  |||||||| | |  | |||| | || |  || | |
Db      8 CTAAAAATGAGATTTCAGATGATGAATGGGAAAATACTAAAACTACGGATGTAAACACAT 67

Qy    878 ctgagttttatg 889
          ||||  ||  ||
Db     68 TTGAGATTGTTG 79


RESULT   7
ROTVP7AN
LOCUS       ROTVP7AN                  87 bp ss-RNA     linear   VRL 23-SEP-1994
DEFINITION  Homo sapiens rotavirus 2 (strain V197) outer capsid protein (VP7)
            RNA, partial cds.
ACCESSION   M37353
VERSION     M37353.1  GI:548274
KEYWORDS    VP7 gene; outer capsid protein.
SOURCE      Human rotavirus 2 (strain V197) RNA.
  ORGANISM  Human rotavirus 2
            Viruses; dsRNA viruses; Reoviridae; Rotavirus.
REFERENCE   1  (bases 1 to 87)
  AUTHORS   Green,K.Y., Sears,J.F., Taniguchi,K., Midthun,K., Hoshino,Y.,
            Gorziglia,M., Nishikawa,K., Urasawa,S., Kapikian,A.Z., Chanock,R.M.
            and Flores,J.
  TITLE     Prediction of human rotavirus serotype by nucleotide sequence
            analysis of the VP7 protein gene
  JOURNAL   J. Virol. 62, 1819-1823 (1988)
  MEDLINE   88188272
FEATURES             Location/Qualifiers
     source          1. .87
                     /organism="Human rotavirus 2"
                     /strain="V197"
                     /db_xref="taxon:36429"
     gene            1. .87
                     /gene="VP7"
     CDS             <1. .>87
                     /gene="VP7"
                     /codon_start=1
```

```
                     /product="outer capsid protein"
                     /protein_id="AAA47372.1"
                     /db_xref="GI:548275"
                     /translation="AEAKNEISDDEWENTKTTDVNTFEIVASS"
BASE COUNT       33 a     11 c     20 g     23 t
ORIGIN


  Query Match             0.6%;  Score 25.6;  DB 14;  Length 87;
  Best Local Similarity   59.7%;  Pred. No. 1.4e+05;
  Matches   43;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy    818 ctgtgaaggtgatttcagtggatgaattgaatgacactatagctgctaatttgagtgaca 877
          ||   || | ||||||||  ||||||| | |  | |||| | || |  || | |
Db      8 CTAAAAATGAGATTTCAGATGATGAATGGGAAAATACTAAAACTACGGATGTAAACACAT 67

Qy    878 ctgagttttatg 889
            |||| ||  ||
Db     68 TTGAGATTGTTG 79


RESULT   8
AF213232
LOCUS       AF213232                  93 bp    DNA     linear   MAM 09-FEB-2001
DEFINITION  Bison bison clone 13.13.1 microsatellite BMS410-93.
ACCESSION   AF213232
VERSION     AF213232.1  GI:11837835
KEYWORDS    .
SOURCE      American bison.
  ORGANISM  Bison bison
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bison.
REFERENCE   1  (bases 1 to 93)
  AUTHORS   Schnabel,R.D., Ward,T.J. and Derr,J.N.
  TITLE     Validation of 15 microsatellites for parentage testing in North
            American bison, Bison bison and domestic cattle
  JOURNAL   Anim. Genet. 31 (6), 360-366 (2000)
  MEDLINE   21070477
REFERENCE   2  (bases 1 to 93)
  AUTHORS   Schnabel,R.D., Ward,T.J. and Derr,J.N.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-1999) Veterinary Pathobiology, Texas A&M
            University, Bldg. 1197 Rm.226, College Station, TX 77843-4467, USA
FEATURES             Location/Qualifiers
     source          1. .93
                     /organism="Bison bison"
                     /db_xref="taxon:9901"
                     /clone="13.13.1"
     primer_bind     1. .22
     repeat_region   1. .93
                     /note="microsatellite BMS410-93"
                     /rpt_type=tandem
     primer_bind     74. .93
BASE COUNT       31 a     35 c      8 g     19 t
ORIGIN


  Query Match             0.6%;  Score 25.4;  DB 4;  Length 93;
  Best Local Similarity   58.7%;  Pred. No. 1.6e+05;
  Matches   44;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy   1460 tgctggctgagccaccacgaatcctcacacctgcaaacacactctaccaggtcattgcaa 1519
          | ||  |  |||   |||  |  | |||||   || ||||||  |   || ||  ||  ||
Db     15 TTCTTTCATAGCACGCACACACACACACACACACACACACACACACACACGTGGTTATAA 74

Qy   1520 acaggcctgctttac 1534
          |||   | ||||| |
Db     75 ACACCACAGCTTTTC 89


RESULT   9
AF272967/c
LOCUS       AF272967                  94 bp    DNA     linear   VRT 08-JAN-2002
DEFINITION  Mastacembelus aculeatus doublesex-like protein Dmrt2a gene, partial
            cds.
ACCESSION   AF272967
VERSION     AF272967.1  GI:10567711
KEYWORDS    .
SOURCE      Mastacembelus aculeatus.
  ORGANISM  Mastacembelus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Synbranchiformes;
            Mastacembelidae; Mastacembelus.
REFERENCE   1  (bases 1 to 94)
  AUTHORS   Huang,X. and Zhou,R.
  TITLE     A conserved family of doublesex related genes from fish
  JOURNAL   J. Exp. Zool. (2002) In press
REFERENCE   2  (bases 1 to 94)
  AUTHORS   Huang,X. and Zhou,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2000) Genetics, Wuhan University, Luojia, Wuhan,
            Hubei 430072, P.R. China
FEATURES             Location/Qualifiers
     source          1. .94
                     /organism="Mastacembelus aculeatus"
                     /db_xref="taxon:138129"
     mRNA            <1. .>94
                     /product="doublesex-like protein Dmrt2a"
     CDS             <1. .>94
                     /note="contains DM domain"
                     /codon_start=2
                     /product="doublesex-like protein Dmrt2a"
                     /protein_id="AAG18552.1"
                     /db_xref="GI:10567712"
                     /translation="VVSCLKGHKRFCRWRDCQCANCLLVVERQRV"
BASE COUNT       13 a     23 c     39 g     19 t
ORIGIN


  Query Match             0.6%;  Score 25.4;  DB 5;  Length 94;
  Best Local Similarity   61.2%;  Pred. No. 1.6e+05;
  Matches   41;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy    166 cccctgattctcttcctgtgccagatgattagtgcactggaagtacctcttgatccaaaa 225
          || |||  ||||  ||     |||    |||| ||||||| ||| ||||  |   || ||
Db     91 CCGCTGCCTCTCCACCACCAGCAGGCAGTTAGCGCACTGGCAGTCCCTCCAGCGGCAGAA 32

Qy    226 cttcttg 232
             ||||
Db     31 GCGCTTG 25


RESULT  10
I71468
LOCUS       I71468                    70 bp    DNA     linear   PAT 03-APR-1998
DEFINITION  Sequence 9 from patent US 5681942.
ACCESSION   I71468
VERSION     I71468.1  GI:3007603
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   Buchwald,M., Strathdee,C.A., Wevrick,R. and Mathew,C.George.Porter.
  TITLE     Fanconi Anemia Type C gene
  JOURNAL   Patent: US 5681942-A 9 28-OCT-1997;
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="unknown"
BASE COUNT       26 a      8 c     11 g     25 t
ORIGIN


  Query Match             0.6%;  Score 25;  DB 6;  Length 70;
```

```
  Best Local Similarity   61.5%;  Pred. No. 2e+05;
  Matches   40;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   3449 aaattgtaaatggttctcggagcttctttgggttaaagggtctaatgccaggaacagcat 3508
          | |  ||||||  | ||| |  ||||||||  |||||  |  || |   | |  | || |
Db      4 AGAGAGTAAATCTTGCTCTGCACTTCTTTGAATTAAATTGATTATTTAAAAGTGCTGCTT 63

Qy   3509 acaaa 3513
          | |||
Db     64 AAAAA 68


RESULT  11
AF174509/c
LOCUS       AF174509                 97 bp    DNA     linear   VRT 20-APR-2000
DEFINITION  Torrentophryne aspinia KIZ-93A011 cytochrome b gene, partial cds;
            mitochondrial gene for mitochondrial product.
ACCESSION   AF174509
VERSION     AF174509.1  GI:7620455
KEYWORDS    .
SOURCE      Torrentophryne aspinia.
  ORGANISM  Mitochondrion Torrentophryne aspinia
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae;
            Torrentophryne.
REFERENCE   1  (bases 1 to 97)
  AUTHORS   Liu,W., Lathrop,A., Fu,J., Yang,D. and Murphy,R.W.
  TITLE     Phylogeny of East Asian bufonids inferred from mitochondrial DNA
            sequences (Anura: Amphibia)
  JOURNAL   Mol. Phylogenet. Evol. 14 (3), 423-435 (2000)
  MEDLINE   20179527
   PUBMED   10712847
REFERENCE   2  (bases 1 to 97)
  AUTHORS   Liu,W., Lathrop,A., Fu,J. and Murphy,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-1999) CBCB, Royal Ontario Museum, 100 Queen's
            Park, Toronto, Ontario M5S 2C6, Canada
FEATURES             Location/Qualifiers
     source          1. .97
                     /organism="Torrentophryne aspinia"
                     /organelle="mitochondrion"
                     /specimen_voucher="KIZ-93A011"
                     /db_xref="taxon:103606"
                     /country="China: Yunnan, Yangbi Co."
     mRNA            <1. .>97
                     /product="cytochrome b"
     CDS             <1. .97
                     /codon_start=2
                     /transl_table=2
                     /product="cytochrome b"
                     /protein_id="AAF64628.1"
                     /db_xref="GI:7620456"
                     /translation="MIGQLASVSYFLIFIIFTPLLGLTENKLTQL"
BASE COUNT       28 a     31 c      8 g     30 t
ORIGIN


  Query Match             0.6%;  Score 24.8;  DB 5;  Length 97;
  Best Local Similarity   54.3%;  Pred. No. 2.4e+05;
  Matches   50;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy   3908 agttcaatgaggatggctcctttattggacaatacagtggtaagaaagagaaagagccgg 3967
          ||||  | | ||  ||  | |  |       |  |  | ||||| || || |||||    |
Db     93 AGTTGAGTTAGTTTGTTTTCGGTGAGTCCGAGAAGGGGGGTAAAAATGATAAAGATTAAG 34

Qy   3968 ctgaaggaaacgaaagctcagaggcaccttct 3999
          | |||| ||| | ||| |  |||   || |
Db     33 AAGTAGGAGACGGAGGCTAATTGGCCAATTAT 2


RESULT  12
AF174510/c
LOCUS       AF174510                 97 bp    DNA     linear   VRT 20-APR-2000
DEFINITION  Torrentophryne tuberospinia KIZ-91A089 cytochrome b gene, partial
            cds; mitochondrial gene for mitochondrial product.
ACCESSION   AF174510
VERSION     AF174510.1  GI:7620457
KEYWORDS    .
SOURCE      Torrentophryne tuberospinia.
  ORGANISM  Mitochondrion Torrentophryne tuberospinia
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Neobatrachia; Bufonoidea; Bufonidae;
            Torrentophryne.
REFERENCE   1  (bases 1 to 97)
  AUTHORS   Liu,W., Lathrop,A., Fu,J., Yang,D. and Murphy,R.W.
  TITLE     Phylogeny of East Asian bufonids inferred from mitochondrial DNA
            sequences (Anura: Amphibia)
  JOURNAL   Mol. Phylogenet. Evol. 14 (3), 423-435 (2000)
  MEDLINE   20179527
   PUBMED   10712847
REFERENCE   2  (bases 1 to 97)
  AUTHORS   Liu,W., Lathrop,A., Fu,J. and Murphy,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUL-1999) CBCB, Royal Ontario Museum, 100 Queen's
            Park, Toronto, Ontario M5S 2C6, Canada
FEATURES             Location/Qualifiers
     source          1. .97
                     /organism="Torrentophryne tuberospinia"
                     /organelle="mitochondrion"
                     /specimen_voucher="KIZ-91A089"
                     /db_xref="taxon:103607"
                     /country="China: Yunnan, Tengchong Co."
     mRNA            <1. .>97
                     /product="cytochrome b"
     CDS             <1. .97
                     /codon_start=2
                     /transl_table=2
                     /product="cytochrome b"
                     /protein_id="AAF64629.1"
                     /db_xref="GI:7620458"
                     /translation="MIGQLASVSYFLIFIIFTPLLGLTENKLTQL"
BASE COUNT       29 a     30 c      8 g     30 t
ORIGIN


  Query Match             0.6%;  Score 24.8;  DB 5;  Length 97;
  Best Local Similarity   54.3%;  Pred. No. 2.4e+05;
  Matches   50;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy   3908 agttcaatgaggatggctcctttattggacaatacagtggtaagaaagagaaagagccgg 3967
          ||||  | | ||  ||  | |  |       |  |  | ||||| || || |||||    |
Db     93 AGTTGAGTTAGTTTGTTTTCTGTGAGTCCGAGAAGGGGGGTGAAAATGATAAAGATTAAG 34

Qy   3968 ctgaaggaaacgaaagctcagaggcaccttct 3999
          | |||| ||| ||||| |  | |   || |
Db     33 AAGTAGGAGACGGAAGCTAATTGACCGATTAT 2


RESULT  13
SYNDEL951/c
LOCUS       SYNDEL951                90 bp    DNA     linear   SYN 27-APR-1993
DEFINITION  X.laevis DNA containing insertion element Tn9, deleted element, 5'
            end.
ACCESSION   M10780
VERSION     M10780.1  GI:208208
KEYWORDS    deletion; deletion mutant.
SEGMENT     1 of 2
SOURCE      X.laevis oocyte 5S DNA and transposon Tn9 DNA, combined in plasmid
            pMB9, mutants 95 and 10.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (bases 1 to 90)
  AUTHORS   Fedoroff,N.V.
  TITLE     Structure of deletion derivatives of a recombinant plasmid
            containing the transposable element Tn9 in the spacer sequence of
            Xenopus laevis 5S DNA
```

```
  JOURNAL   Cold Spring Harb. Symp. Quant. Biol. 43, 1287-1292 (1979)
  MEDLINE   80023367
FEATURES             Location/Qualifiers
     source          1. .90
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
BASE COUNT       28 a      9 c      8 g     45 t
ORIGIN


  Query Match             0.6%;  Score 24.6;  DB 12;  Length 90;
  Best Local Similarity   55.2%;  Pred. No. 2.7e+05;
  Matches   48;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy   3089 ttaacagcacacatgaattaggccctctggtagatttgaaaattcctgccaacaagacac 3148
          || | | |    |  || |    || ||  |  |||||||| ||   |    || ||
Db     90 TTGAAAACTTTGAAAAAATTTAAACTTTGAAAACTTTGAAAAATCAAACTTTGAAAACGT 31

Qy   3149 ggtggactttaaaaaatttaaatttca 3175
          |   ||||| ||||||| || ||| |
Db     30 TGAAAACTTTGAAAAATTAAACTTTGA 4


RESULT  14
AF326227
LOCUS       AF326227                 70 bp    DNA     linear   PLN 02-MAR-2001
DEFINITION  Arabidopsis thaliana small nucleolar RNA R83, complete sequence.
ACCESSION   AF326227
VERSION     AF326227.1  GI:13183695
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 70)
  AUTHORS   Barneche,F., Gaspin,C., Guyot,R. and Echeverria,M.
  TITLE     Extensive characterization of methylation guide small nucleolar
            RNAs in Arabidopsis thaliana
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 70)
  AUTHORS   Barneche,F., Gaspin,C., Guyot,R. and Echeverria,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2000) Laboratoire Genome et Developement des
            Plantes, Universite de Perpignan, UMR CNRS 5096, Avenue de
            Villeneuve, Perpignan, Pyrenees Orientales 66860, France
FEATURES             Location/Qualifiers
     source          1. .70
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /dev_stage="seedling"
                     /note="ecotype: Columbia 0"
     misc_RNA        1. .70
                     /note="AtsnoR83"
                     /product="small nucleolar RNA R83"
                     /evidence=experimental
BASE COUNT       19 a      9 c     19 g     23 t
ORIGIN


  Query Match             0.6%;  Score 24.4;  DB 8;  Length 70;
  Best Local Similarity   63.8%;  Pred. No. 2.9e+05;
  Matches   37;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy   3764 aagatgatgggacatttggagaatacagtgatgcagaagaccacaagcctttgaaaaa 3821
          | |||||||||   |||||||| ||| ||| |  | | | | |   | | |||  ||
Db      5 ATGATGATGGGGATTTTGGAGATTACGATGAGGATGTACAGCTCCCTCTTCTGATTAA 62


RESULT  15
S45981
LOCUS       S45981                   86 bp    RNA     linear   BCT 07-MAY-1993
DEFINITION  Selenoprotein A [Clostridium purinolyticum, Genomic RNA, 86 nt].
ACCESSION   S45981
VERSION     S45981.1  GI:1679975
KEYWORDS    .
SOURCE      Clostridium purinilyticum.
  ORGANISM  Clostridium purinilyticum
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1  (bases 1 to 86)
  AUTHORS   Garcia,G.E. and Stadtman,T.C.
  TITLE     Selenoprotein a component of the glycine reductase complex from
            Clostridium purinolyticum: nucleotide sequence of the gene shows
            that selenocysteine is encoded by UGA
  JOURNAL   J. Bacteriol. 173 (15), 4908 (1991)
  MEDLINE   91310603
  REMARK    GenBank staff at the National Library of Medicine created this
            entry [NCBI gibbsq 45981] from the original journal article.
            This sequence comes from 5a.
COMMENT     On Nov 21, 1996 this sequence version replaced gi:1619721.
            **ERRATUM**  Vol. 173, no. 6, p. 2097.
FEATURES             Location/Qualifiers
     source          1. .86
                     /organism="Clostridium purinilyticum"
                     /db_xref="taxon:1503"
     gene            1. .86
                     /gene="Selenoprotein A"
BASE COUNT       27 a      14 c     24 g     21 t
ORIGIN


  Query Match             0.6%;  Score 24.4;  DB 1;  Length 86;
  Best Local Similarity   56.1%;  Pred. No. 3e+05;
  Matches   46;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy   1130 aatgtatagcaaaaaatgcattaggagccatccaccataccatttctgttagagttaaag 1189
          ||||| | |    || ||||  |||||| ||  | |.    ||       | ||||||||
Db      3 AATGTTTTGTCTGAACTGCAGCAGGAGCTATGGATCTGGAGATCCAACAAAAAGTTAAAG 62

Qy   1190 cggctccatactggatcacagc 1211
            ||| |  | |  |||  |||
Db     63 ATGCTGCTGAGTCAATCGGAGC 84


Search completed: August 13, 2002, 22:40:53
Job time: 11355 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 20:24:26 ; Search time 416.67 Seconds
(without alignments)
17034.402 Million cell updates/sec

Title: US-09-301-380-1
Perfect score: 4134
Sequence: 1 cttcaaagttccccgcatga.....tgggtagagagtatattttc 4134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 1183566

Minimum DB seq length: 15
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	91	2.2	91	22	AAK51341	Human bone marrow
c	2	91	2.2	91	22	AAI57408	Probe #26094 used
	3	50	1.2	51	22	AAL26826	Human SNP oligonuc
	4	48.8	1.2	60	21	AAZ38172	Human Nr-CAM gene
	5	38	0.9	38	21	AAZ38153	Human Nr-CAM gene
	6	28.8	0.7	97	21	AAC14618	Human secreted pro
c	7	28	0.7	29	19	AAV40897	Probe for coding s
c	8	28	0.7	29	21	AAA64165	Probe used to isol
c	9	28	0.7	29	22	AAF98540	Human cDNA clone C
c	10	26.4	0.6	50	21	AAZ38180	Human Nr-CAM gene
c	11	26	0.6	82	17	AAT44495	Murine p57KIP2 gen
c	12	26	0.6	96	22	ABA36168	Probe #14634 for g
c	13	26	0.6	96	22	AAI49423	Probe #18109 used
	14	25.8	0.6	71	13	AAQ24813	Universal promoter
	15	25.8	0.6	73	13	AAQ24815	Universal promoter
c	16	25.4	0.6	88	21	AAA28515	Primer PRIL 5' for
c	17	25.2	0.6	72	22	AAF61467	MBP/BMP fusion con
c	18	25	0.6	39	21	AAZ38169	Antisense Nr-CAM 1
	19	25	0.6	65	13	AAQ24812	Universal promoter
	20	25	0.6	67	13	AAQ24814	Universal promoter
	21	25	0.6	70	14	AAQ51433	Human FACC intron
	22	25	0.6	93	16	AAT22832	Human gene signatu
	23	24.8	0.6	86	16	AAT19434	Human gene signatu
	24	24.8	0.6	93	22	ABA50916	Human breast cell
	25	24.8	0.6	93	22	ABA68894	Human foetal liver
	26	24.8	0.6	93	22	ABA35845	Probe #14311 for g
	27	24.8	0.6	93	22	AAK17226	Human brain expres
	28	24.8	0.6	93	22	AAK43012	Human bone marrow
	29	24.8	0.6	93	22	AAI23781	Probe #13714 for g
	30	24.8	0.6	93	22	AAI49091	Probe #17777 used
	31	24.8	0.6	93	22	AAI09387	Probe #9378 used t
c	32	24.6	0.6	50	21	AAC25209	Human secreted pro
	33	24.4	0.6	38	21	AAZ38156	Human Nr-CAM gene
	34	24.4	0.6	38	21	AAZ38170	Antisense Nr-CAM 2
c	35	24.2	0.6	84	18	AAT71750	Fibrobacter succin
	36	24.2	0.6	95	22	AAS04616	Gene expression pr
	37	24	0.6	24	21	AAZ38159	Human Nr-CAM gene
c	38	24	0.6	38	21	AAZ38157	Human Nr-CAM gene
c	39	24	0.6	38	21	AAZ38171	Antisense Nr-CAM 2
	40	24	0.6	92	21	AAA77898	cDNA encoding huma
	41	24	0.6	92	22	AAI28636	Colon tumour relat
c	42	24	0.6	100	24	AAS98937	Human prostate can
	43	23.8	0.6	37	21	AAZ38168	Antisense Nr-CAM 1
	44	23.8	0.6	48	21	AAZ38179	Human Nr-CAM gene
c	45	23.8	0.6	87	19	AAT97234	Trimmed enzyme sig

ALIGNMENTS

RESULT 1
AAK51341/c
ID AAK51341 standard; DNA; 91 BP.
XX
AC AAK51341;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 25898.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 25898; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 91 BP; 24 A; 23 C; 23 G; 21 T; 0 other;

Query Match 2.2%; Score 91; DB 22; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 atgccgaaaaagaagcgcttatctgcgggcagagtgcccctgattctcttcctgtgccag 189
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 91 ATGCCGAAAAAGAAGCGCTTATCTGCGGGCAGAGTGCCCCTGATTCTCTTCCTGTGCCAG 32

Qy 190 atgattagtgcactggaagtacctcttgatc 220
|||||||||||||||||||||||||||||||
Db 31 ATGATTAGTGCACTGGAAGTACCTCTTGATC 1

RESULT 2
AAI57408/c
ID AAI57408 standard; DNA; 91 BP.
XX
AC AAI57408;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #26094 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 26094; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 91 BP; 24 A; 23 C; 23 G; 21 T; 0 other;

Query Match 2.2%; Score 91; DB 22; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.5e-16;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 atgccgaaaaagaagcgcttatctgcgggcagagtgcccctgattctcttcctgtgccag 189
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 91 ATGCCGAAAAAGAAGCGCTTATCTGCGGGCAGAGTGCCCCTGATTCTCTTCCTGTGCCAG 32

Qy 190 atgattagtgcactggaagtacctcttgatc 220
|||||||||||||||||||||||||||||||
Db 31 ATGATTAGTGCACTGGAAGTACCTCTTGATC 1

RESULT 3
AAL26826
ID AAL26826 standard; DNA; 51 BP.
XX
AC AAL26826;
XX
DT 24-JAN-2002 (first entry)
XX
DE Human SNP oligonucleotide #34.
XX
KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
KW amyloid protein; angiopoietin; apoptosis related protein; cadherin;
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
KW complement related protein; cytochrome; kinesin; cytokine; interferon;
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
KW multifactorial disease; autoimmune disease; infection;
KW nervous system disease; ss.
XX
OS Homo sapiens.
XX
PN WO200147944-A2.
XX
PD 05-JUL-2001.
XX
PF 28-DEC-2000; 2000WO-US35498.
XX
PR 28-DEC-1999; 99US-0173419.
PR 27-DEC-2000; 2000US-0173419.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2001-465210/50.
XX
PT Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT oncogenes and histones, useful for diagnosing and treating, e.g.
PT cancer, autoimmune diseases and infections -
XX
PS Claim 1; Page 1409; 4143pp; English.
XX
CC The present invention relates to oligonucleotides encoding polymorphic
CC variants of proteins related to amylases, amyloid proteins, angiopoietin,
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC histones, kinases, colony stimulating factors, complement related
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC G-protein coupled receptors and thioesterases. The present sequence is
CC one such oligonucleotide. The oligonucleotides and the peptides encoded
CC by them may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate expression of the proteins listed
CC above. Disorders that may be prevented, diagnosed and/or treated include
CC multifactorial diseases with a genetic component, such as autoimmune

```
CC   diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC   systemic lupus erythromatosus and Grave's disease), inflammation, cancer
CC   (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC   leukaemia), diseases of the nervous system and an infection of pathogenic
CC   organisms.
XX
SQ   Sequence 51 BP; 14 A; 13 C; 11 G; 12 T; 1 other;


  Query Match             1.2%;  Score 50;  DB 22;  Length 51;
  Best Local Similarity   98.0%;  Pred. No. 0.00014;
  Matches   50;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1177 gttagagttaaagcggctccatactggatcacagcccctcaaaatcttgtg 1227
           ||||||||||||||||||||||| |||||||||||||||||||||||||
Db      1 gttagagttaaagcggctccatacnggatcacagcccctcaaaatcttgtg 51


RESULT   4
AAZ38172
ID   AAZ38172 standard; DNA; 60 BP.
XX
AC   AAZ38172;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Human Nr-CAM gene fragment.
XX
KW   Nr-CAM; neuron-glia-related cell adhesion molecule; cell proliferation;
KW   tumorigenesis; malignancy; cancer; leukemia; hyperproliferative disorder;
KW   degenerative disorder; growth deficiency; trauma; wound; tumor; diabetes;
KW   systemic lupus erythematosus; demyelinating disease; growth; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO9955380-A1.
XX
PD   04-NOV-1999.
XX
PF   27-APR-1999;   99WO-US09039.
XX
PR   27-APR-1998;   98US-0083152.
PR   14-DEC-1998;   98US-0112098.
XX
PA   (PACI-) PACIFIC NORTHWEST CANCER FOUND.
XX
PI   Murphy GP,  Boynton AL,  Sehgal A;
XX
DR   WPI; 2000-023268/02.
XX
PT   Use of neuron-glia-related cell adhesion molecule for developing agents
PT   for the diagnosis and treatment of e.g. cancers, hyperproliferative
PT   disorders, growth deficiencies, degenerative disorders, trauma or
PT   wounds  -
XX
PS   Examples; Page 123; 183pp; English.
XX
CC   The invention relates to the use of neuron-glia-related cell adhesion
CC   molecule (Nr-CAM) as a marker for diagnosing, treating, inhibiting or
CC   preventing disorders involving cell proliferation. An antisense nucleic
CC   acid complementary to at least a portion of an RNA transcript of a
CC   Nr-CAM gene can be used to inhibit hyperproliferation of a tumor cell,
CC   for the treatment of tumorigenesis. Agents which inhibit and promote
CC   Nr-CAM function can be used for the treatment of various diseases and
CC   disorders (see AAZ38152 for a detailed description). The products can
CC   also be used for detection, diagnosis and production of animal models.
CC   The present sequence represents a human Nr-CAM gene fragment against
CC   which phosphorothioate oligonucleotides are designed.
XX
SQ   Sequence 60 BP; 22 A; 9 C; 16 G; 13 T; 0 other;


  Query Match             1.2%;  Score 48.8;  DB 21;  Length 60;
  Best Local Similarity   88.3%;  Pred. No. 0.00034;
  Matches   53;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy     99 agcaggagttaagctaatgcagcttaaaataatgccgaaaaagaagcgcttatctgcggg 158
          || ||     | ||||||||||||||||||||||||||||||||||||||||||||||
Db      1 aggagttaagatgctaatgcagcttaaaataatgccgaaaaagaagcgcttatctgcggg 60


RESULT   5
AAZ38153
ID   AAZ38153 standard; DNA; 38 BP.
XX
AC   AAZ38153;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Human Nr-CAM gene fragment (basepairs 4097-4134).
XX
KW   Nr-CAM; neuron-glia-related cell adhesion molecule; cell proliferation;
KW   tumorigenesis; malignancy; cancer; leukemia; hyperproliferative disorder;
KW   degenerative disorder; growth deficiency; trauma; wound; tumor; diabetes;
KW   systemic lupus erythematosus; demyelinating disease; growth; human; ss.
XX
OS   Homo sapiens.
XX
PN   WO9955380-A1.
XX
PD   04-NOV-1999.
XX
PF   27-APR-1999;   99WO-US09039.
XX
PR   27-APR-1998;   98US-0083152.
PR   14-DEC-1998;   98US-0112098.
XX
PA   (PACI-) PACIFIC NORTHWEST CANCER FOUND.
XX
PI   Murphy GP,  Boynton AL,  Sehgal A;
XX
DR   WPI; 2000-023268/02.
XX
PT   Use of neuron-glia-related cell adhesion molecule for developing agents
PT   for the diagnosis and treatment of e.g. cancers, hyperproliferative
PT   disorders, growth deficiencies, degenerative disorders, trauma or
PT   wounds  -
XX
PS   Disclosure; Fig 2C; 183pp; English.
XX
CC   The invention relates to the use of neuron-glia-related cell adhesion
CC   molecule (Nr-CAM) as a marker for diagnosing, treating, inhibiting or
CC   preventing disorders involving cell proliferation. An antisense nucleic
CC   acid complementary to at least a portion of an RNA transcript of a
CC   Nr-CAM gene can be used to inhibit hyperproliferation of a tumor cell,
CC   for the treatment of tumorigenesis. Agents which inhibit Nr-CAM function
CC   can be used to treat or prevent malignancies, e.g. brain cancer,
CC   leukemia, B cell lymphoma, premalignant conditions, benign tumors,
CC   hyperproliferative disorders or benign dysproliferative disorders. Such
CC   treatment is especially useful for treating glioblastoma, glioma,
CC   meningioma, astrocytoma, medulloblastoma, neuroectodermal cancer and
CC   neuroblastoma, especially glioblastoma multiforme. Agents which promote
CC   Nr-CAM function may also be used to treat or prevent degenerative
CC   disorders, growth deficiencies, hypoproliferative disorders, physical
CC   trauma, lesions or wounds. In particular, they can be used for treating
CC   e.g. traumatic, ischemic, malignant and degenerative lesions and
CC   alcoholic cerebellar degeneration. Also described is the treatment of
CC   lesions associated with systemic diseases e.g. diabetes or systemic lupus
CC   erythematosus, lesions caused by toxic substances e.g. alcohol, lead or
CC   other toxins; and demyelinated lesions of the nervous system, in which a
CC   portion of the nervous system is destroyed or injured by a demyelinating
CC   disease e.g. multiple sclerosis, HIV-associated myelopathy, transverse
CC   myelopathy of various etiologies, progressive multifocal
CC   leukoencephalopathy or central pontine myelinolysis; or lesions of the
CC   central or peripheral nervous systems. In addition, agents which promote
CC   Nr-CAM function can be promoted to increase growth of animals (e.g. cows,
```

```
CC   horses, pigs, goats, deer, chickens) and plants (particularly edible
CC   plants, e.g. tomatoes, melons, lettuce, carrots, potatoes, and other
CC   vegetables), particularly those that are food or material sources. They
CC   can also be used in vitro e.g. to expand cells e.g. stem cells,
CC   progenitor cells, muscle cells, fibroblasts, or liver cells to grow
CC   cells/tissue in vitro prior to administration to a patient. The products
CC   can also be used for detection, diagnosis and production of animal
CC   models. The present sequence represents a previously cloned human Nr-CAM
CC   gene (accn no: AAU55258) fragment.
XX
SQ   Sequence 38 BP; 12 A; 5 C; 7 G; 14 T; 0 other;


  Query Match             0.9%;  Score 38;  DB 21;  Length 38;
  Best Local Similarity   100.0%;  Pred. No. 0.4;
  Matches   38;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   4097 tctcatactatgaacatatgggtagagagtatattttc 4134
         |||||||||||||||||||||||||||||||||||||||
Db      1 tctcatactatgaacatatgggtagagagtatattttc 38


RESULT   6
AAC14618
ID   AAC14618 standard; cDNA; 97 BP.
XX
AC   AAC14618;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 18693.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 18693; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 97 BP; 43 A; 8 C; 28 G; 14 T; 4 other;


  Query Match             0.7%;  Score 28.8;  DB 21;  Length 97;
  Best Local Similarity   60.0%;  Pred. No. 3.8e+02;
  Matches   42;  Conservative    3;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   3853 aaaaaagaagatagtgacgacagcctagttgactatggagaaggggttaatggccagttc 3912
         ||  ||||||| |   | |||||||| || ||  ||| |::|| |    : |  |    |
Db     21 aatgaagaagaaaacaaagacagccttgtagatgatgaasmagagaaasmagatcttggc 80

Qy   3913 aatgaggatg 3922
           ||||||||||
Db     81 gatgaggatg 90


RESULT   7
AAV40897/c
ID   AAV40897 standard; DNA; 29 BP.
XX
AC   AAV40897;
XX
DT   25-SEP-1998  (first entry)
XX
DE   Probe for coding sequence of clone CO722_1.
XX
KW   Human; nutritional supplement; cell proliferation/differentiation;
KW   cytokine; immunostimulant; immunosuppressant; haematopoiesis regulator;
KW   receptor/ligand activity; cadherin/tumour invasion suppressor; probe;
KW   anti-inflammatory; tumour inhibitor; clone CO722_1; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9824905-A2.
XX
PD   11-JUN-1998.
XX
PF   05-DEC-1997;   97WO-US22211.
XX
PR   03-DEC-1997;   97US-0984516.
PR   06-DEC-1996;   96US-0762216.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1998-333324/29.
XX
PT   New isolated polynucleotides encoding secreted polypeptides -
PT   isolated from a human foetal kidney cDNA library, a human adult
PT   blood cDNA library or a human adult brain cDNA library
XX
PS   Disclosure; Page 93; 109pp; English.
XX
CC   This sequence represents a probe for the coding sequence of clone
CC   CO722_1 of the invention. The clone was isolated from a human adult
CC   brain cDNA library. The DNAs and proteins can be used as nutritional
CC   sources or supplements, or may exhibit cytokine and cell
CC   proliferation/differentiation activity, immune stimulating or suppressing
CC   activity, haematopoiesis regulating activity, receptor/ligand activity,
CC   anti-inflammatory activity, cadherin/tumour invasion suppressor activity,
CC   tumour inhibition activity or other activities.
XX
SQ   Sequence 29 BP; 7 A; 9 C; 4 G; 8 T; 1 other;


  Query Match             0.7%;  Score 28;  DB 19;  Length 29;
  Best Local Similarity   96.6%;  Pred. No. 3e+02;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     75 aaggaaattcagtgtgtgagtctcagcag 103
          |||||||||||||||||||||||||||| |
```

```
Db      29 AAGGAAATTCAGTGTGTGAGTCTCAGCNG 1


RESULT   8
AAA64165/c
ID   AAA64165 standard; DNA; 29 BP.
XX
AC   AAA64165;
XX
DT   20-DEC-2000 (first entry)
XX
DE   Probe used to isolate cDNA encoding secreted protein clone CO722_1.
XX
KW   Human; secreted protein; immune deficiency; infectious disease;
KW   tissue growth; wound healing; tissue repair; burn; incision; ulcer;
KW   osteoporosis; osteoarthritis; periodontal disease; tooth repair;
KW   nervous system disorder; angiogenic activity; fibrosis; fertility;
KW   reperfusion injury; systemic cytokine damage; contraceptive;
KW   thrombolytic; coagulation disorder; antiinflammatory activity;
KW   inflammatory condition; probe; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   modified_base    2
FT                    /*tag= a
FT                    /note= "biotinylated phosphoramidite residue"
XX
PN   WO200050592-A1.
XX
PD   31-AUG-2000.
XX
PF   24-FEB-2000; 2000WO-US04731.
XX
PR   24-FEB-1999;   99US-0256938.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  Lavallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V;
XX
DR   WPI; 2000-558399/51.
XX
PT   New secreted polypeptides and nucleic acids encoding them used for
PT   treating various anemias, bone, cartilage, tendon, ligament and/or
PT   nerve tissue or regeneration, contraceptives, and nutritional
PT   supplements  -
XX
PS   Disclosure; Page 115; 119pp; English.
XX
CC   The present sequence represents a probe for cDNA encoding a human
CC   secreted protein. The polypeptides may be used in the treatment of
CC   various immune deficiencies and disorders, and to treat infectious
CC   diseases caused by viral, bacterial, fungal or other infections. The
CC   polypeptides are also used in compositions for bone, cartilage, tendon,
CC   ligament and/or nerve tissue growth or regeneration, for wound healing
CC   and tissue repair and replacement, and in the treatment of burns,
CC   incisions and ulcers. The protein is used in the treatment of
CC   osteoporosis or osteoarthritis, periodontal disease and other tooth
CC   repair processes, and nervous system disorders. They also exhibit
CC   angiogenic activity and protect, regenerate and treat lung or liver
CC   fibrosis, reperfusion injury in various tissues, and conditions
CC   resulting from systemic cytokine damage. They promote or inhibit tissue
CC   differentiation and are used as contraceptives and to enhance fertility.
CC   They also have a hemostatic or thrombolytic activity and can be used
CC   to treat various coagulation disorders. They also have an
CC   antiinflammatory activity and can be used to treat inflammatory
CC   conditions.
XX
SQ   Sequence 29 BP; 7 A; 9 C; 4 G; 8 T; 1 other;

  Query Match             0.7%;  Score 28;  DB 21;  Length 29;
  Best Local Similarity   96.6%;  Pred. No. 3e+02;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      75 aaggaaattcagtgtgtgagtctcagcag 103
           ||||||||||||||||||||||||||| |
Db      29 AAGGAAATTCAGTGTGTGAGTCTCAGCNG 1


RESULT   9
AAF98540/c
ID   AAF98540 standard; DNA; 29 BP.
XX
AC   AAF98540;
XX
DT   07-JUN-2001 (first entry)
XX
DE   Human cDNA clone CO722_1 specific probe SEQ ID 250.
XX
KW   Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW   differentiation; immune system modulator; tissue growth; chemotactic;
KW   haemostatic; thrombolytic; anti-inflammatory; tumour inhibition; ss;
KW   haematopoiesis; probe.
XX
OS   Homo sapiens.
XX
PN   WO200119988-A1.
XX
PD   22-MAR-2001.
XX
PF   14-SEP-2000; 2000WO-US25135.
XX
PR   17-SEP-1999;   99US-0398829.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Jacobs K,  McCoy JM,  LaVallie ER,  Collins-Racie LA,  Evans C;
PI   Merberg D,  Treacy M,  Bowman MR,  Spaulding V,  Agostino MJ;
XX
DR   WPI; 2001-244801/25.
XX
PT   Isolated nucleic acids encoding polypeptides, useful for modulating
PT   e.g. cytokine and cell proliferation/differentiation activity, the
PT   immune system and hematopoiesis regulating activity -
XX
PS   Disclosure; Page 540; 557pp; English.
XX
CC   Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC   proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC   tissue types, and may be used in the prevention, treatment and diagnosis
CC   of diseases associated with inappropriate protein expression. The
CC   polypeptides and nucleic acids may be used as nutrients or to modulate
CC   cytokine and cell proliferation/differentiation activity and may also be
CC   involved in modulation of the immune system. The cDNA sequences,
CC   proteins, their agonists and/or antagonists exhibit haematopoiesis
CC   regulating activity; tissue growth activity; activin/inhibin activity;
CC   chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC   activity; receptor/ligand activity; anti-inflammatory activity;
CC   haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC   tumour inhibition activity. Included in the invention are probes
CC   represented in AAF98490 - AAF98572 which are specific for the cDNA clones
CC   encoding the secreted proteins.
XX
SQ   Sequence 29 BP; 7 A; 9 C; 4 G; 8 T; 1 other;

  Query Match             0.7%;  Score 28;  DB 22;  Length 29;
  Best Local Similarity   96.6%;  Pred. No. 3e+02;
  Matches   28;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      75 aaggaaattcagtgtgtgagtctcagcag 103
           ||||||||||||||||||||||||||| |
Db      29 AAGGAAATTCAGTGTGTGAGTCTCAGCNG 1
```

```
RESULT  10
AAZ38180/c
ID   AAZ38180 standard; DNA; 50 BP.
XX
AC   AAZ38180;
XX
DT   14-FEB-2000  (first entry)
XX
DE   Human Nr-CAM gene specific primer BT307.
XX
KW   Nr-CAM; neuron-glia-related cell adhesion molecule; cell proliferation;
KW   tumorigenesis; malignancy; cancer; leukemia; hyperproliferative disorder;
KW   degenerative disorder; growth deficiency; trauma; wound; tumor; diabetes;
KW   systemic lupus erythematosus; demyelinating disease; PCR primer; ss.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO9955380-A1.
XX
PD   04-NOV-1999.
XX
PF   27-APR-1999;   99WO-US09039.
XX
PR   27-APR-1998;   98US-0083152.
PR   14-DEC-1998;   98US-0112098.
XX
PA   (PACI-) PACIFIC NORTHWEST CANCER FOUND.
XX
PI   Murphy GP,  Boynton AL,  Sehgal A;
XX
DR   WPI; 2000-023268/02.
XX
PT   Use of neuron-glia-related cell adhesion molecule for developing agents
PT   for the diagnosis and treatment of e.g. cancers, hyperproliferative
PT   disorders, growth deficiencies, degenerative disorders, trauma or
PT   wounds -
XX
PS   Examples; Page 124; 183pp; English.
XX
CC   The invention relates to the use of neuron-glia-related cell adhesion
CC   molecule (Nr-CAM) as a marker for diagnosing, treating, inhibiting or
CC   preventing disorders involving cell proliferation. An antisense nucleic
CC   acid complementary to at least a portion of an RNA transcript of a
CC   Nr-CAM gene can be used to inhibit hyperproliferation of a tumor cell,
CC   for the treatment of tumorigenesis. Agents which inhibit and promote
CC   Nr-CAM function can be used for the treatment of various diseases and
CC   disorders (see AAZ38152 for a detailed description). The products can
CC   also be used for detection, diagnosis and production of animal models.
CC   Sequences AAZ38179-80 represent primers specific for the human Nr-CAM
CC   gene.
XX
SQ   Sequence 50 BP; 15 A; 11 C; 11 G; 13 T; 0 other;


  Query Match             0.6%;  Score 26.4;  DB 21;  Length 50;
  Best Local Similarity   83.3%;  Pred. No. 1.3e+03;
  Matches   30;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy   1399 gtatatcagtgcaatgcctctaatgaatatggatat 1434
            ||||   |   |||||||||||||||||||||||||
Db     50 GTATGGGATCCCAATGCCTCTAATGAATATGGATAT 15


RESULT  11
AAT44495/c
ID   AAT44495 standard; DNA; 82 BP.
XX
AC   AAT44495;
XX
DT   22-FEB-1997  (first entry)
XX
DE   Murine p57KIP2 gene translation initiation region.
XX
KW   Cyclin-dependent kinase inhibitor; CDK; CDI; p57KIP2; cell cycle;
KW   cancer; hyperplasia; ulcer; hyperproliferation; hypoproliferation;
KW   diagnosis; gene therapy; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   exon            1..15
FT                   /*tag=  a
FT                   /note= "3' end of exon sequence"
FT   intron          16..28
FT                   /*tag=  b
FT                   /note= "base n at position 22 denotes an approx.
FT                           200 bp intronic sequence not provided in
FT                           Fig 4"
FT   exon            29..82
FT                   /*tag=  c
FT                   /codon_start= 39
FT                   /note= "there is an alternative codon start site
FT                           at position 78 for use with B and B'
FT                           splicing events"
FT   misc_feature    29
FT                   /*tag=  d
FT                   /label= A
FT                   /note= "p57 acceptor site A"
FT   misc_feature    67
FT                   /*tag=  e
FT                   /label= B
FT                   /note= "alternative acceptor site B"
FT   misc_feature    70
FT                   /*tag=  f
FT                   /label= B'
FT                   /note= "alternative acceptor site B'"
XX
PN   WO9631534-A1.
XX
PD   10-OCT-1996.
XX
PF   03-APR-1996;   96WO-US04563.
XX
PR   03-APR-1995;   95US-0415655.
XX
PA   (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI   Lee M,  Massague J;
XX
DR   WPI; 1996-464971/46.
DR   P-PSDB; AAW01438.
XX
PT   Mammalian p57-KIP2 and related DNA - used to develop prods. useful
PT   for diagnosis and treatment of hyper- and hypo-proliferative
PT   disorders
XX
PS   Disclosure; Fig 4; 98pp; English.
XX
CC   3 Distinct cDNA clones (AAT44496-98) are derived from genomic DNA
CC   (AAT44495) in the translation start region of the murine gene
CC   (see also AAT44494) coding for cyclin-dependent kinase inhibitor
CC   p57KIP2 (AAW01437).  These result from splicing at position A,
CC   leading to the full-length (348-amino acid) protein having the
CC   N-terminal sequence shown in AAW01438, or alternative splicing at
CC   positions B or B', resulting in a 335-amino acid product,
CC   designated p57KIP2B.  The intron sequence in the genomic DNA was
CC   identified by PCR analysis (see also AAT44500-01).
XX
SQ   Sequence 82 BP; 21 A; 27 C; 21 G; 12 T; 1 other;


  Query Match             0.6%;  Score 26;  DB 17;  Length 82;
  Best Local Similarity   58.7%;  Pred. No. 2.3e+03;
  Matches   44;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;
```

```
Qy   1840 tccctcactgtcctgtggctgaaggacaacagggaactgcccagtgatgaaaggttcact 1899
          ||| || |||| |||  || || | |||    |||  ||||||  |  ||| | |   |
Db     81 TCCATCGCTGTTCTGCTGCGGAGGTACACGTCGGACATGCCCATGGTCGAAGGCTGTGCN 22

Qy   1900 gttgacaaggatcat 1914
          |   ||  || | ||
Db     21 GGCTACCTGGCTGAT 7
```

RESULT 12
ABA36168/c
ID ABA36168 standard; DNA; 96 BP.
XX
AC ABA36168;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #14634 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 14634; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 96 BP; 24 A; 21 C; 31 G; 20 T; 0 other;

Query Match 0.6%; Score 26; DB 22; Length 96;
Best Local Similarity 65.5%; Pred. No. 2.6e+03;
Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy   1541 gtgccttctttgggtctcctctcccaaccatcgagtggtttaaaggagctaaaggaag 1598
          || ||| || | ||| |||||| || ||  |  || |||   || ||||   || |||
Db     79 GTTCCTGCTCTTGGTTTCCTCTGCCCACTCTGCAGGGGTAATAATGAGCCCCAGAAAG 22
```

RESULT 13
AAI49423/c
ID AAI49423 standard; DNA; 96 BP.
XX
AC AAI49423;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #18109 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 18109; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 96 BP; 24 A; 21 C; 31 G; 20 T; 0 other;

Query Match 0.6%; Score 26; DB 22; Length 96;
Best Local Similarity 65.5%; Pred. No. 2.6e+03;
Matches 38; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy   1541 gtgccttctttgggtctcctctcccaaccatcgagtggtttaaaggagctaaaggaag 1598
          || ||| || | ||| |||||| || ||  |  || |||   || ||||   || |||
Db     79 GTTCCTGCTCTTGGTTTCCTCTGCCCACTCTGCAGGGGTAATAATGAGCCCCAGAAAG 22
```

RESULT 14
AAQ24813
ID AAQ24813 standard; DNA; 71 BP.
XX
AC AAQ24813;
XX
DT 19-NOV-1992 (first entry)
XX
DE Universal promoter-2.
XX
KW UP; untranslated leader sequence; E-PCR; T7 RNA polymerase; SP6;
KW T3; T4; ghl; alfalfa mosaic virus; ds.

```
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        10..26
FT                   /*tag=  a
FT                   /label= T7_promoter
FT   5'UTR           27..64
FT                   /*tag=  b
FT                   /label= UTL_from_AMV
XX
PN   WO9207949-A.
XX
PD   14-MAY-1992.
XX
PF   05-NOV-1991;   91WO-US08291.
XX
PR   05-NOV-1990;   90US-0609318.
XX
PA   (USSA ) US SEC OF ARMY.
XX
PI   Kain KC,  Lanar DE;
XX
DR   WPI; 1992-183684/22.
XX
PT   Protein in vitro prodn. for efficient transcription - avoids
PT   cloning of deoxyribonucleic acid by expression polymerase chain
PT   reaction for translation
XX
PS   Disclosure; Fig 2; 37pp; English.
XX
CC   The sequences given in AAQ24812 - AAQ24815 are universal promoter (UP)
CC   sequences which contain a double stranded RNA transcription
CC   promoter, an untranslated leader sequence enhancing translation
CC   activity followed by three codons, the first of which encodes Met.
CC   These UP's can be used in a process described as expression
CC   polymerase chain reaction (E-PCR).  The active polymerase binding
CC   site used in this invention is the one for viral T7 RNA polymerase,
CC   however, any promoter site may be used that corresponds to the RNA
CC   polymerase that will be employed for the transcription of the DNA.
CC   Promoter sites for suitable polymerases that could be used are
CC   those for the SP6 polymerase, the T3 or T4 phage polymerases or the
CC   ghl promoter.  The untranslated leader (UTL) sequence between the T7
CC   promoter and the initial ATG codon is derived from the coat protein
CC   mRNA of the alfalfa mosaic virus (AMV).  In vitro translation of mRNA
CC   is often dependant on the presence of, and the characteristics of, an
CC   UTL sequence 5' to the initiation codon.  It has been shown that
CC   replacement of a genes native UTL with the AMV UTL can increase
CC   translational efficiency by as much as 35-fold.  Other UTL sequences
CC   may be substituted for this AMV UTL.
XX
SQ   Sequence 71 BP; 19 A; 16 C; 7 G; 29 T; 0 other;


  Query Match             0.6%;  Score 25.8;  DB 13;  Length 71;
  Best Local Similarity   63.9%;  Pred. No. 2.4e+03;
  Matches   39;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    626 ttggattaccaccacctataatattttggatggataattcctttcaaagacttccacaaa 685
          ||  | ||| ||   |||||   ||||   |    | || ||||||||  |||||||| |
Db      7 ttctaatacgactcactatagggtttttatttttaattttctttcaaatacttccaccat 66

Qy    686 g 686
          |
Db     67 g 67


RESULT  15
AAQ24815
ID   AAQ24815 standard; DNA; 73 BP.
XX
AC   AAQ24815;
XX
DT   19-NOV-1992  (first entry)
XX
DE   Universal promoter-4.
XX
KW   UP; untranslated leader sequence; E-PCR; T7 RNA polymerase; SP6;
KW   T3; T4; ghl; alfalfa mosaic virus; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   promoter        10..26
FT                   /*tag=  a
FT                   /label= T7_promoter
FT   5'UTR           27..64
FT                   /*tag=  b
FT                   /label= UTL_from_AMV
XX
PN   WO9207949-A.
XX
PD   14-MAY-1992.
XX
PF   05-NOV-1991;   91WO-US08291.
XX
PR   05-NOV-1990;   90US-0609318.
XX
PA   (USSA ) US SEC OF ARMY.
XX
PI   Kain KC,  Lanar DE;
XX
DR   WPI; 1992-183684/22.
XX
PT   Protein in vitro prodn. for efficient transcription - avoids
PT   cloning of deoxyribonucleic acid by expression polymerase chain
PT   reaction for translation
XX
PS   Disclosure; Fig 4; 37pp; English.
XX
CC   The sequences given in AAQ24812 - AAQ24815 are universal promoter (UP)
CC   sequences which contain a double stranded RNA transcription
CC   promoter, an untranslated leader sequence enhancing translation
CC   activity followed by three codons, the first of which encodes Met.
CC   These UP's can be used in a process described as expression
CC   polymerase chain reaction (E-PCR).  The active polymerase binding
CC   site used in this invention is the one for viral T7 RNA polymerase,
CC   however, any promoter site may be used that corresponds to the RNA
CC   polymerase that will be employed for the transcription of the DNA.
CC   Promoter sites for suitable polymerases that could be used are
CC   those for the SP6 polymerase, the T3 or T4 phage polymerases or the
CC   ghl promoter.  The untranslated leader (UTL) sequence between the T7
CC   promoter and the initial ATG codon is derived from the coat protein
CC   mRNA of the alfalfa mosaic virus (AMV).  In vitro translation of mRNA
CC   is often dependant on the presence of, and the characteristics of, an
CC   UTL sequence 5' to the initiation codon.  It has been shown that
CC   replacement of a genes native UTL with the AMV UTL can increase
CC   translational efficiency by as much as 35-fold.  Other UTL sequences
CC   may be substituted for this AMV UTL.
XX
SQ   Sequence 73 BP; 19 A; 17 C; 8 G; 29 T; 0 other;


  Query Match             0.6%;  Score 25.8;  DB 13;  Length 73;
  Best Local Similarity   63.9%;  Pred. No. 2.4e+03;
  Matches   39;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy    626 ttggattaccaccacctataatattttggatggataattcctttcaaagacttccacaaa 685
          ||  | ||| ||   |||||   ||||   |    | || ||||||||  |||||||| |
Db      7 ttctaatacgactcactatagggtttttatttttaattttctttcaaatacttccaccat 66

Qy    686 g 686
          |
Db     67 g 67
```

ug 14 08:10:40 2002

Search completed: August 13, 2002, 22:49:51
Job time 8725 sec

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 19:38:44 ; Search time 82.1 Seconds
(without alignments)
12368.440 Million cell updates/sec

Title: US-09-301-380-1
Perfect score: 4134
Sequence: 1 cttcaaagttccccgcatga.....tgggtagagagtatattttc 4134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 576162

Minimum DB seq length: 15
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	25	0.6	70	1	US-08-441-430-9	Sequence 9, Appli
c	2	24	0.6	100	4	US-09-564-805-23	Sequence 23, Appl
c	3	23.6	0.6	67	4	US-09-025-769B-83	Sequence 83, Appl
	4	23.4	0.6	66	2	US-08-709-874A-15	Sequence 15, Appl
	5	23.4	0.6	66	4	US-09-104-382-15	Sequence 15, Appl
c	6	23	0.6	60	3	US-08-415-655-10	Sequence 10, Appl
c	7	22.8	0.6	63	1	US-08-386-495-1	Sequence 1, Appli
c	8	22.8	0.6	63	5	PCT-US96-02331-1	Sequence 1, Appli
c	9	22.8	0.6	78	1	US-08-446-102-1	Sequence 1, Appli
	10	22.8	0.6	78	1	US-08-446-102-2	Sequence 2, Appli
	11	22.8	0.6	78	4	US-08-617-010C-15	Sequence 15, Appl
	12	22.8	0.6	78	4	US-09-566-591-15	Sequence 15, Appl
	13	22.8	0.6	100	4	US-09-242-690A-58	Sequence 58, Appl
	14	22.6	0.5	94	1	US-08-105-483-447	Sequence 447, App
c	15	22.6	0.5	94	1	US-08-105-483-448	Sequence 448, App
	16	22.6	0.5	94	1	US-08-709-209-447	Sequence 447, App
c	17	22.6	0.5	94	1	US-08-709-209-448	Sequence 448, App
	18	22.6	0.5	94	1	US-08-303-275-159	Sequence 159, App
c	19	22.6	0.5	94	1	US-08-303-275-160	Sequence 160, App
	20	22.6	0.5	94	1	US-08-458-101-447	Sequence 447, App
c	21	22.6	0.5	94	1	US-08-458-101-448	Sequence 448, App
c	22	22.6	0.5	100	3	US-08-836-561-61	Sequence 61, Appl
	23	22.4	0.5	78	4	US-09-240-078-43	Sequence 43, Appl
c	24	22.2	0.5	77	1	US-08-242-663A-3	Sequence 3, Appli
c	25	22.2	0.5	77	5	PCT-US95-06132-3	Sequence 3, Appli
c	26	22.2	0.5	79	2	US-08-184-009-158	Sequence 158, App
c	27	22.2	0.5	79	2	US-08-458-356-158	Sequence 158, App
c	28	22.2	0.5	79	4	US-08-460-736-158	Sequence 158, App
	29	22.2	0.5	84	1	US-08-182-175A-27	Sequence 27, Appl
	30	22.2	0.5	84	1	US-08-474-633A-68	Sequence 68, Appl
	31	22.2	0.5	84	5	PCT-US92-06412-27	Sequence 27, Appl
	32	22.2	0.5	95	2	US-08-332-766A-41	Sequence 41, Appl
	33	22	0.5	60	1	US-08-153-848-10	Sequence 10, Appl
	34	22	0.5	60	3	US-09-299-843A-10	Sequence 10, Appl
	35	22	0.5	60	4	US-09-088-337B-10	Sequence 10, Appl
	36	22	0.5	60	5	PCT-US93-11153-10	Sequence 10, Appl
	37	22	0.5	90	1	US-08-513-846-2	Sequence 2, Appli
	38	22	0.5	90	1	US-08-513-846-18	Sequence 18, Appl
	39	22	0.5	93	2	US-08-105-989-12	Sequence 12, Appl
	40	22	0.5	93	3	US-09-138-922-12	Sequence 12, Appl
c	41	21.8	0.5	56	1	US-08-441-430-25	Sequence 25, Appl
	42	21.8	0.5	76	3	US-08-284-516C-22	Sequence 22, Appl
c	43	21.8	0.5	81	1	US-08-238-863-42	Sequence 42, Appl
c	44	21.8	0.5	81	1	US-08-443-407-42	Sequence 42, Appl
c	45	21.8	0.5	81	5	PCT-US95-05600-186	Sequence 186, App

ALIGNMENTS

```
RESULT   1
US-08-441-430-9
; Sequence 9, Application US/08441430
; Patent No. 5681942
;  GENERAL INFORMATION:
;    APPLICANT:  Buchwald, Manuel
;    APPLICANT:  Strathdee, Craig A.
;    APPLICANT:  Wevrick, Rachel
;    APPLICANT:  Mathew, Christopher George Porter
;    TITLE OF INVENTION:  Fanconi Anemia Type C Gene
;    NUMBER OF SEQUENCES:  73
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Richard J. Polley, Esq.
;      ADDRESSEE:  Klarquist, Sparkman, Campbell, Leigh &
;      ADDRESSEE:  Whinston, LLP
;      STREET:  121 S.W. Salmon, Suite 1600
;      CITY:  Portland
;      STATE:  Oregon
;      COUNTRY:  U.S.A.
;      ZIP:  97204
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Disk, 3+-inch
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  MS DOS
;      SOFTWARE:  WordPerfect 5.1/ASCII Text File
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,430
;      FILING DATE:  May 15, 1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. 07/876,285
;      FILING DATE:  April 29, 1992
;      APPLICATION NUMBER:  U.S. 07/918,313
;      FILING DATE:  July 21, 1992
;      APPLICATION NUMBER:  U.S. 08/003,963
;      FILING DATE:  January 15, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Richard J. Polley, Esq.
;      REGISTRATION NUMBER:  28,107
;      REFERENCE/DOCKET NUMBER:  3812-42824
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (503) 226-7391
;      TELEFAX:  (503) 228-9446
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  70 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Double stranded
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Genomic DNA
```

```
;    HYPOTHETICAL:  No
;    ANTI-SENSE:  No
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-08-441-430-9


  Query Match              0.6%;  Score 25;  DB 1;  Length 70;
  Best Local Similarity   61.5%;  Pred. No. 5.2e+02;
  Matches   40;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   3449 aaattgtaaatggttctcggagcttctttgggttaaagggtctaatgccaggaacagcat 3508
          | |  ||||||  | ||| |  |||||||| |||||  |  || |   | |  | || |
Db      4 AGAGAGTAAATCTTGCTCTGCACTTCTTTGAATTAAATTGATTATTTAAAAGTGCTGCTT 63

Qy   3509 acaaa 3513
          | |||
Db     64 AAAAA 68


RESULT   2
US-09-564-805-23/c
; Sequence 23, Application US/09564805
; Patent No. 6333403
; GENERAL INFORMATION:
;  APPLICANT:  Tavtigian, Sean V.
;  APPLICANT:  Teng, David H.F.
;  APPLICANT:  Simard, Jacques
;  APPLICANT:  Rommens, Johanna M.
;  APPLICANT:  Myriad Genetics, Inc.
;  TITLE OF INVENTION:  Chromosome 17p-Linked Prostate Cancer Susceptibility
;  TITLE OF INVENTION:  Gene and a Paralog and Orthologous Genes
;  FILE REFERENCE:  2318-258
;  CURRENT APPLICATION NUMBER: US/09/564,805
;  CURRENT FILING DATE:  2000-05-05
;  PRIOR APPLICATION NUMBER:  US 60/107,468
;  PRIOR FILING DATE:  1998-11-06
;  PRIOR APPLICATION NUMBER:  09/434,382
;  PRIOR FILING DATE:  1999-11-05
;  NUMBER OF SEQ ID NOS:  240
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 23
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(100)
;   OTHER INFORMATION: exon 20
US-09-564-805-23


  Query Match              0.6%;  Score 24;  DB 4;  Length 100;
  Best Local Similarity   54.5%;  Pred. No. 1.3e+03;
  Matches   48;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy   3297 caaagttcaagctgtaaataccaggatcagcaatcttactgctgcagctgctgagaccta 3356
          ||||   || | |   || | |    | | ||||||| |  ||||||     |||| ||
Db     94 CAAATCACATGTTCGCAACAGCGAACTGATCAATCTTTCCACTGCAGGACTGGAGATCTC 35

Qy   3357 tgccaatatcagttgggaatatgaggga 3384
            |||  |  | |  || | |  |  |||
Db     34 AGCCCCTTCCTGAAGGCATTTGGCAGGA 7


RESULT   3
US-09-025-769B-83/c
; Sequence 83, Application US/09025769B
; Patent No. 6300064
;  GENERAL INFORMATION:
;    APPLICANT:  Knappik, Achim
;    APPLICANT:  Pack, Peter
;    APPLICANT:  Ilag, Vic
;    APPLICANT:  Ge, Liming
;    APPLICANT:  Moroney, Simon
;    APPLICANT:  Plueckthun, Andreas
;    TITLE OF INVENTION:  Protein/(Poly)peptide libraries
;    NUMBER OF SEQUENCES:  373
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  James F. Haley, Jr., Esq. c/o Fish & Neave
;      STREET:  1251 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10021
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/025,769B
;      FILING DATE:  18-FEB-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95 11 3021.0
;      FILING DATE:  18-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  James F. Haley, Jr., Esq.
;      REGISTRATION NUMBER:  27,794
;      REFERENCE/DOCKET NUMBER:  MORPHO/5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)596-9000
;      TELEFAX:  (212)596-9090
;  INFORMATION FOR SEQ ID NO:  83:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  67 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
;      DESCRIPTION:  /desc = "synthetic oligonucleotide"
US-09-025-769B-83


  Query Match              0.6%;  Score 23.6;  DB 4;  Length 67;
  Best Local Similarity   76.3%;  Pred. No. 1.3e+03;
  Matches   29;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy   1228 ctgtccccaggagaggatgggaccttgatctgcagagc 1265
          ||||| || || ||   || |||| ||| |||||||||
Db     39 CTGTCTCCGGGCGAACGTGCGACCCTGAGCTGCAGAGC 2


RESULT   4
US-08-709-874A-15
; Sequence 15, Application US/08709874A
; Patent No. 5854040
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Chiba, Shigeru
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Trans-4-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS, LLP
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/709,874A
;       FILING DATE:  09-SEP-1996
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/301,653
;       FILING DATE:  07-SEP-1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/482,554
;       FILING DATE:  07-JUN-1995
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Terry, David T.
;       REGISTRATION NUMBER:  20178
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  703-312-6600
;       TELEFAX:  703-312-6666
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  66 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid, synthetic DNA
US-08-709-874A-15


  Query Match             0.6%;  Score 23.4;  DB 2;  Length 66;
  Best Local Similarity   60.0%;  Pred. No. 1.5e+03;
  Matches   39;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy   3266 ctggtattcttccacctgatgtaggtgcaggcaaagttcaagctgtaaataccaggatca 3325
          | || |||| || || | | |  ||  ||||| |  ||||| | | | ||| |  | |
Db      1 CCGGAATTCGTCGACTTCACGCGGGCCCAGGCCATCTTCAATCAGCAGATAGCCCGCTTC 60

Qy   3326 gcaat 3330
          | ||
Db     61 ACGAT 65


RESULT   5
US-09-104-382-15
; Sequence 15, Application US/09104382
; Patent No. 6242231
;  GENERAL INFORMATION:
;    APPLICANT:  Ozaki, Akio
;    APPLICANT:  Mori, Hideo
;    APPLICANT:  Shibasaki, Takeshi
;    APPLICANT:  Ando, Katsuhiko
;    APPLICANT:  Chiba, Shigeru
;    TITLE OF INVENTION:  Process for Producing
;    TITLE OF INVENTION:  Trans-4-Hydroxy-L-Proline
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ANTONELLI, TERRY, STOUT AND KRAUS, LLP
;      STREET:  1300 NORTH SEVENTEENTH STREET
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  U.S.A.
;      ZIP:  22209
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/104,382
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/709,874
;      FILING DATE:  09-SEP-1996
;      APPLICATION NUMBER:  08/301,653
;      FILING DATE:  07-SEP-1994
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/482,554
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terry, David T.
;      REGISTRATION NUMBER:  20178
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-312-6600
;      TELEFAX:  703-312-6666
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  66 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid, synthetic DNA
US-09-104-382-15


  Query Match             0.6%;  Score 23.4;  DB 4;  Length 66;
  Best Local Similarity   60.0%;  Pred. No. 1.5e+03;
  Matches   39;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy   3266 ctggtattcttccacctgatgtaggtgcaggcaaagttcaagctgtaaataccaggatca 3325
          | || |||| || || | | |  ||  ||||| |  ||||| | | | ||| |  | |
Db      1 CCGGAATTCGTCGACTTCACGCGGGCCCAGGCCATCTTCAATCAGCAGATAGCCCGCTTC 60

Qy   3326 gcaat 3330
          | ||
Db     61 ACGAT 65


RESULT   6
US-08-415-655-10/c
; Sequence 10, Application US/08415655
; Patent No. 6025480
;  GENERAL INFORMATION:
;    APPLICANT:  Massague, Joan
;    APPLICANT:  Lee, Mong-hong
;    TITLE OF INVENTION:  ISOLATED NUCLEIC ACID MOLECULES ENCODING
;    TITLE OF INVENTION:  p75KIP2, A CYCLIN-DEPENDENT KINASE INHIBITOR AND USES OF
;    TITLE OF INVENTION:  SAME
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  United States of America
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/415,655
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  1747/47418
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  10:
```

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  60 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-415-655-10


  Query Match             0.6%;  Score 23;  DB 3;  Length 60;
  Best Local Similarity   63.6%;  Pred. No. 1.8e+03;
  Matches   35;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1840 tccctcactgtcctgtggctgaaggacaacagggaactgcccagtgatgaaaggt 1894
          ||| || |||| |||  || || | |||    |||  ||||||  |  ||| | |
Db     59 TCCATCGCTGTTCTGCTGCGGAGGTACACGTCGGACATGCCCATGGTCGAAGGCT 5


RESULT   7
US-08-386-495-1/c
; Sequence 1, Application US/08386495
; Patent No. 5753434
;  GENERAL INFORMATION:
;    APPLICANT:  Ryner, Lisa C.
;    APPLICANT:  Baker, Bruce S.
;    APPLICANT:  Wasserman, Steven A.
;    APPLICANT:  Castrillon, Diego H.
;    TITLE OF INVENTION:  Methods and Compositions for Altering
;    TITLE OF INVENTION:  Sexual Behavior
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/386,495
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8600-0153
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  63 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  3x repeat probe
US-08-386-495-1


  Query Match             0.6%;  Score 22.8;  DB 1;  Length 63;
  Best Local Similarity   66.0%;  Pred. No. 2.2e+03;
  Matches   33;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy   1608 tgaagatatttatgttttacatgaaaatggaactttggaaatcaaagatg 1657
          |||||||   | | | ||   |||| ||||| || ||   ||  ||||||
Db     54 TGAAGATGGATCTATGTTGATTGAAGATGGATCTATGTTGATTGAAGATG 5


RESULT   8
PCT-US96-02331-1/c
; Sequence 1, Application PC/TUS9602331
;  GENERAL INFORMATION:
;    APPLICANT:  The Board of Trustees of the Leland Stanford Junior
;    APPLICANT:  University
;    APPLICANT:  Board of Reagents, The University of Texas System
;    TITLE OF INVENTION:  Methods and Compositions for Altering
;    TITLE OF INVENTION:  Sexual Behavior
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dehlinger & Associates
;      STREET:  350 Cambridge Avenue, Suite 250
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94306
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/02331
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/386,495
;      FILING DATE:  10-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sholtz, Charles K.
;      REGISTRATION NUMBER:  38,615
;      REFERENCE/DOCKET NUMBER:  8600-0153.41
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 324-0880
;      TELEFAX:  (415) 324-0960
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  63 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      INDIVIDUAL ISOLATE:  3x repeat probe
PCT-US96-02331-1


  Query Match             0.6%;  Score 22.8;  DB 5;  Length 63;
  Best Local Similarity   66.0%;  Pred. No. 2.2e+03;
  Matches   33;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy   1608 tgaagatatttatgttttacatgaaaatggaactttggaaatcaaagatg 1657
          |||||||   | | | ||   |||| ||||| || ||   ||  ||||||
Db     54 TGAAGATGGATCTATGTTGATTGAAGATGGATCTATGTTGATTGAAGATG 5


RESULT   9
US-08-446-102-1/c
; Sequence 1, Application US/08446102
; Patent No. 5753439
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, Cassandra L.
;    APPLICANT:  YAAR, Ron
;    APPLICANT:  SZAFRANSKI, Przemyslaw
;    APPLICANT:  CANTOR, Charles R.
```

```
;    TITLE OF INVENTION:  NUCLEIC ACID DETECTION METHODS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BAKER & BOTTS, L.L.P.
;      STREET:  1299 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004-2400
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,102
;      FILING DATE:  19-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Remenick, James
;      REGISTRATION NUMBER:  36,902
;      REFERENCE/DOCKET NUMBER:  16865.0199
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-639-7700
;      TELEFAX:  202-639-7890
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  78 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-446-102-1


  Query Match             0.6%;  Score 22.8;  DB 1;  Length 78;
  Best Local Similarity   59.1%;  Pred. No. 2.5e+03;
  Matches   39;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   1994 ccagcgctgtgcttagcgttgttgctcctactccaactccagctcccgtttacgatgtcc 2053
          |||| | |  | |||||||   |||| || || |  || | |||      | ||| |
Db     69 CCAGGGATTCGGTTAGCGTGACTGCTGCTGCTGCTGCTGCTGGATGATCCGACGCAT 10

Qy   2054 caaatc 2059
          || |||
Db      9 CAGATC 4


RESULT  10
US-08-446-102-2
; Sequence 2, Application US/08446102
; Patent No. 5753439
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, Cassandra L.
;    APPLICANT:  YAAR, Ron
;    APPLICANT:  SZAFRANSKI, Przemyslaw
;    APPLICANT:  CANTOR, Charles R.
;    TITLE OF INVENTION:  NUCLEIC ACID DETECTION METHODS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BAKER & BOTTS, L.L.P.
;      STREET:  1299 Pennsylvania Avenue, N.W.
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20004-2400
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/446,102
;      FILING DATE:  19-MAR-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Remenick, James
;      REGISTRATION NUMBER:  36,902
;      REFERENCE/DOCKET NUMBER:  16865.0199
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-639-7700
;      TELEFAX:  202-639-7890
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  78 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-446-102-2


  Query Match             0.6%;  Score 22.8;  DB 1;  Length 78;
  Best Local Similarity   59.1%;  Pred. No. 2.5e+03;
  Matches   39;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   1994 ccagcgctgtgcttagcgttgttgctcctactccaactccagctcccgtttacgatgtcc 2053
          |||| | |  | |||||||   |||| || || |  || | |||      | ||| |
Db     10 CCAGGGATTCGGTTAGCGTGACTGCTGCTGCTGCTGCTGCTGGATGATCCGACGCAT 69

Qy   2054 caaatc 2059
          || |||
Db     70 CAGATC 75


RESULT  11
US-08-617-010C-15
; Sequence 15, Application US/08617010C
; Patent No. 6194144
;  GENERAL INFORMATION:
;    APPLICANT:  Hubert K ster
;    TITLE OF INVENTION:  DNA SEQUENCING BY MASS SPECTROMETRY
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  4250 Executive Square, 7th Floor
;      CITY:  La Jolla
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92037-9103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  ASCII
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/617,010C
;      FILING DATE:  18-MAR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
```

```
;       APPLICATION NUMBER:  08/178,216
;       FILING DATE:  06-JAN-1994
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/001,323
;       FILING DATE:  07-JAN-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Seidman, Stephanie L
;       REGISTRATION NUMBER:  33,779
;       REFERENCE/DOCKET NUMBER:  24736-2012
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  858-450-8400
;       TELEFAX:  619-587-5360
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  15:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  78 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-617-010C-15


  Query Match             0.6%;  Score 22.8;  DB 4;  Length 78;
  Best Local Similarity   59.1%;  Pred. No. 2.5e+03;
  Matches   39;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   1994 ccagcgctgtgcttagcgttgttgctcctactccaactccagctcccgtttacgatgtcc 2053
          |||| | |  | |||||||  |||| || || |  || | |||      | ||| |
Db     10 CCAGGGATTCGGTTAGCGTGACTGCTGCTGCTGCTGCTGCTGGATGATCCGACGCAT 69

Qy   2054 caaatc 2059
          || |||
Db     70 CAGATC 75


RESULT  12
US-09-566-591-15
; Sequence 15, Application US/09566591
; Patent No. 6238871
;   GENERAL INFORMATION:
;       APPLICANT: Hubert K"ster
;       TITLE OF INVENTION: DNA SEQUENCING BY MASS SPECTROMETRY
;       NUMBER OF SEQUENCES: 21
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Heller Ehrman White & McAuliffe
;            STREET: 4250 Executive Square, 7th Floor
;            CITY: La Jolla
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 92037-9103
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette
;            COMPUTER: IBM Compatible
;            OPERATING SYSTEM: DOS
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/566,591
;            FILING DATE: 08-May-2000
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/617,010
;            FILING DATE: 18-MAR-1996
;            APPLICATION NUMBER: 08/178,216
;            FILING DATE: 06-JAN-1994
;            APPLICATION NUMBER: 08/001,323
;            FILING DATE: 07-JAN-1993
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Seidman, Stephanie L
;            REGISTRATION NUMBER: 33,779
;            REFERENCE/DOCKET NUMBER: 24736-2012B
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 858-450-8400
;            TELEFAX: 858-587-5360
;            TELEX: <Unknown>
;   INFORMATION FOR SEQ ID NO: 15:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 78 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-566-591-15


  Query Match             0.6%;  Score 22.8;  DB 4;  Length 78;
  Best Local Similarity   59.1%;  Pred. No. 2.5e+03;
  Matches   39;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   1994 ccagcgctgtgcttagcgttgttgctcctactccaactccagctcccgtttacgatgtcc 2053
          |||| | |  | |||||||  |||| || || |  || | |||      | ||| |
Db     10 CCAGGGATTCGGTTAGCGTGACTGCTGCTGCTGCTGCTGCTGGATGATCCGACGCAT 69

Qy   2054 caaatc 2059
          || |||
Db     70 CAGATC 75


RESULT  13
US-09-242-690A-58
; Sequence 58, Application US/09242690A
; Patent No. 6284534
; GENERAL INFORMATION:
;  APPLICANT: KONDO, KEIJI
;  APPLICANT:  MIURA, YUTAKA
;  TITLE OF INVENTION: YEAST VECTOR AND METHOD OF PRODUCING PROTEINS USING THE
;  TITLE OF INVENTION:  SAME
;  FILE REFERENCE: 049441/0118
;  CURRENT APPLICATION NUMBER: US/09/242,690A
;  CURRENT FILING DATE:  1999-02-23
;  PRIOR APPLICATION NUMBER: PCT/JP97/02924
;  PRIOR FILING DATE: 1997-08-22
;  PRIOR APPLICATION NUMBER: JP 8/241062
;  PRIOR FILING DATE: 1996-08-23
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 58
;   LENGTH: 100
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Primer
US-09-242-690A-58


  Query Match             0.6%;  Score 22.8;  DB 4;  Length 100;
  Best Local Similarity   56.8%;  Pred. No. 2.9e+03;
  Matches   42;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy   2519 atgtatccaaatatattgtctcaggcacgccaacctttgttccatacctgatcaaagttc 2578
          | ||   | |   || ||  ||   ||| |  |||| |||| |  || |||||   |||
Db      1 aagtcgacgagcttaatgattctctcacccttacctctgttaccgacttgatcgtggttt 60

Qy   2579 aggccctgaatgac 2592
            ||   | || |||
Db     61 tggatgtaaacgac 74


RESULT  14
US-08-105-483-447
; Sequence 447, Application US/08105483
; Patent No. 5494807
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
```

```
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/105,483
;      FILING DATE:  12-AUG-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  447:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  94 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-105-483-447


  Query Match             0.5%;  Score 22.6;  DB 1;  Length 94;
  Best Local Similarity   54.1%;  Pred. No. 3.2e+03;
  Matches   46;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     619 cccccaattggattaccaccacctataatattttggatggataattcctttcaaagactt 678
           || ||  || ||  ||    |  ||| || | || |  |||| | ||  ||  ||   | |
Db       4 CCTCCTTTTAGAAAACACGAAGATATTATTTCTTTGTGGGATCAATCTTTAAAACCTCCT 63

Qy     679 ccacaaagtgagagagtttctcaag 703
               ||  || |||||   | |||
Db      64 TTTAGAAAAGATAGAGTAATTGAAG 88


RESULT  15
US-08-105-483-448/c
; Sequence 448, Application US/08105483
; Patent No. 5494807
;  GENERAL INFORMATION:
;    APPLICANT:  Paoletti, Enzo
;    TITLE OF INVENTION:  GENETICALLY ENGINEERED VACCINE
;    TITLE OF INVENTION:  STRAIN
;    NUMBER OF SEQUENCES:  462
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis, Morris & Safford
;      ADDRESSEE:  c/o William S. Frommer
;      STREET:  530 Fifth Avenue
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/105,483
;      FILING DATE:  12-AUG-1993
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/847,951
;      FILING DATE:  06-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Frommer, William S.
;      REGISTRATION NUMBER:  25,506
;      REFERENCE/DOCKET NUMBER:  454310-2400
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 840-3333
;      TELEFAX:  (212) 840-0712
;  INFORMATION FOR SEQ ID NO:  448:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  94 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-105-483-448


  Query Match             0.5%;  Score 22.6;  DB 1;  Length 94;
  Best Local Similarity   54.1%;  Pred. No. 3.2e+03;
  Matches   46;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     619 cccccaattggattaccaccacctataatattttggatggataattcctttcaaagactt 678
           || ||  || ||  ||    |  ||| || | || |  |||| | ||  ||  ||   | |
Db      91 CCTCCTTTTAGAAAACACGAAGATATTATTTCTTTGTGGGATCAATCTTTAAAACCTCCT 32

Qy     679 ccacaaagtgagagagtttctcaag 703
               ||  || |||||   | |||
Db      31 TTTAGAAAAGATAGAGTAATTGAAG 7


Search completed: August 13, 2002, 22:42:33
Job time: 11029 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: August 13, 2002, 19:30:38 ; Search time 2981.25 Seconds
(without alignments)
18715.778 Million cell updates/sec

Title: US-09-301-380-1
Perfect score: 4134
Sequence: 1 cttcaaagttccccgcatga.....tgggtagagagtatattttc 4134

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 297160

Minimum DB seq length: 15
Maximum DB seq length: 100

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	28.8	0.7	85	9	AI096387	AI096387 qb91f10.x
2	27.6	0.7	88	9	AI007269	AI007269 ua33h03.r
3	26.6	0.6	84	10	BE322578	BE322578 NF006E05I
c 4	26.4	0.6	100	9	AW101101	AW101101 sd74a07.y
5	26	0.6	71	10	D78209	D78209 D78209 EST
c 6	26	0.6	83	12	TA293D12Q	AL485648 T. brucei
c 7	26	0.6	92	9	AA470870	AA470870 ne19g03.s
c 8	26	0.6	93	10	BG796297	BG796297 UTSW_SM41
9	26	0.6	95	10	BE324593	BE324593 NF014E05P
c 10	26	0.6	96	10	W03279	W03279 za36c11.r1
c 11	26	0.6	100	9	AW845547	AW845547 MR1-CT005
c 12	26	0.6	100	10	D45290	D45290 HUMHG2360 H
13	25.8	0.6	71	12	AZ826437	AZ826437 2M0102K08
14	25.8	0.6	85	10	BE289162	BE289162 601092758
15	25.8	0.6	99	12	FR0022001	AL014872 F.rubripe
c 16	25.8	0.6	100	10	BF092347	BF092347 MR4-TN010
c 17	25.4	0.6	95	10	BI654803	BI654803 603281144
c 18	25.2	0.6	98	9	AA716301	AA716301 zh28e12.s
19	25.2	0.6	100	9	AV001666	AV001666 AV001666
c 20	25	0.6	87	12	FR0010708	AL001980 F.rubripe
21	25	0.6	100	10	BG221902	BG221902 RST41719
22	24.8	0.6	64	10	BE636385	BE636385 SWOvAMCAQ
c 23	24.8	0.6	73	10	W88076	W88076 mf67e10.r1
c 24	24.8	0.6	80	9	AW246447	AW246447 2821633.3
25	24.8	0.6	84	9	AW675386	AW675386 bb45g09.y
26	24.8	0.6	94	10	R51717	R51717 yg77f06.r1
c 27	24.8	0.6	95	9	AI153902	AI153902 ud50e04.r
c 28	24.8	0.6	96	12	AF166318	AF166318 AF166318
29	24.8	0.6	99	10	BF378532	BF378532 QV1-UM003
30	24.6	0.6	79	9	AA670839	AA670839 vs90e06.r
31	24.6	0.6	89	12	AZ835091	AZ835091 2M0129G11
32	24.6	0.6	97	9	AI628325	AI628325 ts36f01.x
33	24.6	0.6	97	9	AA232334	AA232334 zr27b04.r
34	24.6	0.6	97	9	AW395938	AW395938 sh07f05.y
c 35	24.4	0.6	71	12	AZ336443	AZ336443 1M0066A14
c 36	24.4	0.6	73	10	BM053420	BM053420 id66a07.y
c 37	24.4	0.6	78	9	AA935490	AA935490 nz53c02.s
38	24.4	0.6	85	9	AI477927	AI477927 fb49b10.x
39	24.4	0.6	88	9	AA523959	AA523959 ng26g01.s
c 40	24.4	0.6	94	9	AW496789	AW496789 1tg Neuro
41	24.4	0.6	95	12	AZ468419	AZ468419 1M0281K19
c 42	24.4	0.6	100	9	AA237743	AA237743 mx77a06.r
43	24.2	0.6	64	9	AW599941	AW599941 SWOvAFCAP
44	24.2	0.6	87	10	BI493713	BI493713 df104f06.
c 45	24.2	0.6	88	12	AZ784747	AZ784747 2M0027B09

ALIGNMENTS

RESULT 1
AI096387
LOCUS AI096387 85 bp mRNA linear EST 26-OCT-1998
DEFINITION qb91f10.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:1707499 3' similar to SW:FSA_BOVIN P50291 FOLLISTATIN PRECURSOR ;, mRNA sequence.
ACCESSION AI096387
VERSION AI096387.1 GI:3446298
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 85)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Insert Length: 1428 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.
FEATURES Location/Qualifiers
source 1..85
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1707499"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."

BASE COUNT 16 a 21 c 25 g 23 t
ORIGIN

Query Match 0.7%; Score 28.8; DB 9; Length 85;
Best Local Similarity 65.6%; Pred. No. 8.3e+03;
Matches 42; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy    141 gaagcgcttatctgcgggcagagtgcccctgattctcttcctgtgccagatgattagtgc 200
          || |||| |||||| |||| |   |  ||| | ||  | |||||||||| ||||    |
Db      1 GAGGCGCCTATCTGTGGGCTTATGCCTTCTGCTGCTGCTACTGTGCCAGCTGATGCAGGA 60

Qy    201 actg 204
          |||
Db     61 TCTG 64
```

RESULT 2
AI007269
LOCUS AI007269 88 bp mRNA linear EST 12-JUN-1998
DEFINITION ua33h03.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1348565 5' similar to TR:Q15170 Q15170 ;, mRNA sequence.
ACCESSION AI007269
VERSION AI007269.1 GI:3216826
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 88)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:697357
Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 6.

FEATURES Location/Qualifiers
source 1. .88
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1348565"
/clone_lib="Soares_mammary_gland_NbMMG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT

T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 34 a 10 c 29 g 15 t
ORIGIN

Query Match 0.7%; Score 27.6; DB 9; Length 88;
Best Local Similarity 60.8%; Pred. No. 1.8e+04;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy   1865 acaacagggaactgcccagtgatgaaaggttcactgttgacaaggatcatctagtggtag 1924
          |||||||| |||||  |||||||||   ||| |  | |||  |||| |     | | |
Db     12 ACAACAGGCAACTGAGCAGTGATGATCTGTTTAGAGATGAGGAGGAGCTAGATGAGATCA 71

Qy   1925 ctgatgtcagtgac 1938
          | || ||  ||
Db     72 GGAAAGTGAGAAAC 85
```

RESULT 3
BE322578
LOCUS BE322578 84 bp mRNA linear EST 21-DEC-2000
DEFINITION NF006E05IN1F1036 Insect herbivory Medicago truncatula cDNA clone NF006E05IN 5', mRNA sequence.
ACCESSION BE322578
VERSION BE322578.2 GI:11962897
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 84)
AUTHORS Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula insect herbivory library
JOURNAL Unpublished (2000)
COMMENT On Jul 14, 2000 this sequence version replaced gi:9196355.
Contact: Korth K
Dept. of Plant Pathology
University of Arkansas
217 Plant Science Building, Fayetteville, AR 72701, USA
Tel: 501 575 5191
Fax: 501 575 7601
Email: kkorth@comp.uark.edu
Medicago Genome Initiative accession: MGI:S:25317
Insert Length: 855 Std Error: 0.00
Plate: 006 row: E column: 05
Seq primer: TCACACAGGAAACAGCTATGAC.

FEATURES Location/Qualifiers
source 1. .84
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF006E05IN"
/clone_lib="Insect herbivory"
/tissue_type="local and systemic leaves"
/dev_stage="mature"
/note="Vector: Lambda Zap; Library was produced from fully expanded M. truncatula leaves of plants fed upon by Spodoptera exigua (beet armyworm) for 24 hours. Systemic (undamaged leaves from injured plants) and wounded leaves were harvested and pooled."

BASE COUNT 33 a 10 c 25 g 16 t
ORIGIN

Query Match 0.6%; Score 26.6; DB 10; Length 84;

```
  Best Local Similarity   71.4%;  Pred. No. 3.3e+04;
  Matches   35;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   3410 aatatggtgtagcaggcagcaaagaagaatggagaaaagaaattgtaaa 3458
          ||| ||    || ||| || |||||||||| | ||| ||||||  ||  ||
Db      3 AATCTGACAGAGAAGGAAGGAAAGAAGAAGGTAGAGAAGAAAATGGGAA 51


RESULT   4
AW101101/c
LOCUS       AW101101                 100 bp    mRNA    linear   EST 30-NOV-2001
DEFINITION  sd74a07.y1 Gm-c1008 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1008-2029 5', mRNA sequence.
ACCESSION   AW101101
VERSION     AW101101.1  GI:6071714
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Putative full length read
            vector to vector length is 101 This clone is available through:
            ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL
            35801 For further information call: (800)-533-4363 or contact via
            email: ccu@resgen.com
            Seq primer: -40RP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1008-2029"
                     /clone_lib="Gm-c1008"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     cDNA library was constructed from mRNA isolated from whole
                     young pods, approximately 2cm long, of 12-week-old
                     greenhouse grown plants. The library was prepared using
                     the Life Technologies pSuperScript cDNA library
                     construction kit. Complementary DNA was synthesized from
                     mRNA using a poly (dT) sequence with a NotI restriction
                     site. SalI linkers adapters were ligated to the
                     blunt-ended cDNA fragments followed by NotI digestion. The
                     cDNA fragments were directionally cloned into the
                     NotI-SalI restriction site of the pSPORT1 vector. The
                     ligated cDNA fragments were transformed into E. coli
                     ElectroMax DH10B host cells. This library was constructed
                     by Dr. Lila Vodkin and Dr. Anu Khanna."
BASE COUNT       28 a     12 c     16 g     44 t
ORIGIN


  Query Match             0.6%;  Score 26.4;  DB 9;  Length 100;
  Best Local Similarity   57.1%;  Pred. No. 4.1e+04;
  Matches   48;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy    710 atggggacctttatttttccaatgtcctcccagaggacacccgcgaagactatatctgtt 769
          ||  |||| | | | |    |||||||   || |  | |  ||  |||| |||  | ||
Db     96 ATCAGGACTTATTTATAAAAAATGTCCAATCATAACAAAAACGTAAAGATTATCCCGGTA 37

Qy    770 atgctagatttaatcatactcaaa 793
          ||| || |  |  | |||  || |
Db     36 ATGATATACATTCTTATAGACATA 13


RESULT   5
D78209
LOCUS       D78209                   71 bp    mRNA    linear   EST 23-MAR-1998
DEFINITION  D78209 EST from 8p21.3-p22 Homo sapiens cDNA clone B6-1-5, mRNA
            sequence.
ACCESSION   D78209
VERSION     D78209.1  GI:2104127
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 71)
  AUTHORS   Chinen,K., Isomura,M., Izawa,K., Fujiwara,Y., Ohata,H., Iwamasa,T.
            and Nakamura,Y.
  TITLE     Isolation of 45 exon-like fragments from 8p22-->p21.3, a region
            that is commonly deleted in hepatocellular, colorectal, and
            non-small cell lung carcinomas
  JOURNAL   Cytogenet. Cell Genet. 75 (2-3), 190-196 (1996)
  MEDLINE   97193198
COMMENT     Contact: Yusuke Nakamura
            Institute of Medical Science
            University of Tokyo
            4-6-1, Shirokanedai, Minato-ku, Tokyo 108,  Japan
            Tel: 81-3-5449-5372
            Fax: 81-3-5449-5433
            Email: yusuke@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .71
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="8p21.3-p22"
                     /clone="B6-1-5"
                     /clone_lib="EST from 8p21.3-p22"
BASE COUNT       29 a      8 c     21 g     13 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 10;  Length 71;
  Best Local Similarity   62.1%;  Pred. No. 4.4e+04;
  Matches   41;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy   3840 agacaggactgtgaaaaaagaagatagtgacgacagcctagttgactatggagaaggggt 3899
          ||| || || || |||| ||||| | ||  || ||||   | |  | | ||| |||
Db      5 AGAAAGTACAGTAAAAAGAGAAGGTGGTTTCGGCAGCACTGATAGCAAAGGAAAAGCAAC 64

Qy   3900 taatgg 3905
          | || |
Db     65 TTATTG 70


RESULT   6
TA293D12Q/c
LOCUS       TA293D12Q                83 bp    DNA     linear   GSS 13-DEC-2000
DEFINITION  T. brucei sheared genomic DNA clone 293d12, reverse sequence,
            genomic survey sequence.
ACCESSION   AL485648
VERSION     AL485648.1  GI:11853124
KEYWORDS    GSS.
SOURCE      Trypanosoma brucei.
  ORGANISM  Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE   1  (bases 1 to 83)
```

```
  AUTHORS   Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R.,
            Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
            Melville,S.E., Rajandream,M.A. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nh1@sanger.ac.uk
COMMENT     Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb).  The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects.  In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/Projects/T_brucei/.
FEATURES              Location/Qualifiers
     source           1. .83
                      /organism="Trypanosoma brucei"
                      /strain="TREU927"
                      /db_xref="taxon:5691"
                      /clone="293d12"
BASE COUNT       21 a     18 c      7 g     37 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 12;  Length 83;
  Best Local Similarity   65.5%;  Pred. No. 4.8e+04;
  Matches   38;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy   1349 tagatggcgataccattatttttttcaaatgttcaagaaagatcaagtgcagtatatca 1406
           ||| ||  |  | |||| |||||| |||     |||||| |  ||| | || ||| ||
Db     83 TAGGTGAAGTCAACATTTTTTTTTTGAAAAAAAAAAGAAAAAAAAAGAGGAGAATAGCA 26


RESULT   7
AA470870/c
LOCUS       AA470870                 92 bp    mRNA    linear   EST 14-AUG-1997
DEFINITION  ne19g03.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:881716 3'
            similar to gb:S66915_cds1 ATP SYNTHASE GAMMA CHAIN, MITOCHONDRIAL
            PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   AA470870
VERSION     AA470870.1  GI:2198179
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
            , Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Seq primer: -41m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES              Location/Qualifiers
     source           1. .92
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:881716"
                      /clone_lib="NCI_CGAP_Co3"
                      /sex="pooled"
                      /tissue_type="colon"
                      /lab_host="DH10B"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                      was prepared from 12 pooled bulk tumor samples and primed
                      with a Not I - oligo(dT) primer. Double-stranded cDNA was
                      ligated to Eco RI adaptors (Pharmacia), digested with Not
                      I and cloned into the Not I and Eco RI sites of the
                      modified pT7T3 vector. Library went through one round of
                      normalization. "
BASE COUNT       26 a     20 c      26 g     20 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 9;  Length 92;
  Best Local Similarity   59.5%;  Pred. No. 5.1e+04;
  Matches   44;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy   1268 atggcaaccccaaacccagaattagctggttaacaaatggagtcccaatagaaattgccc 1327
           ||  ||||||  ||||    |         || ||||| ||||| | | | ||||||  |
Db     80 ATTCCAACCCGTAACCGCCCAGGCTGGTCTTCACAAAAGGAGTTCGATTGGAAATTATCT 21

Qy   1328 ctgatgaccccagc 1341
           |||    | |||||
Db     20 CTGGGTCCTCCAGC 7


RESULT   8
BG796297/c
LOCUS       BG796297                 93 bp    mRNA    linear   EST 16-MAY-2001
DEFINITION  UTSW_SM41H9 UTSW Adult Mouse Skeletal Muscle Library Mus musculus
            cDNA clone UTSW_SM41H9, mRNA sequence.
ACCESSION   BG796297
VERSION     BG796297.1  GI:14131867
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 93)
  AUTHORS   Gallardo,T.D., Schageman,J.J., Pertsemlidis,A., Garner,H.R.,
            Williams,R.S. and Shohet,R.V.
  TITLE     UT Southwestern Medical Center, Adult Mouse Skeletal Muscle cDNA
            Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Schageman JJ
            Shohet/Garner Labs
            University of Texas Southwestern Medical Center
            6000 Harry Hines Blvd., NA2.226, Dallas, TX 75390, USA
            Tel: 214 648 1674
            Email: Jeff.Schageman@UTSouthwestern.edu
            cDNA library constructed by UTSW as a component of the Program for
            Genomic Applications (PGA) and the Reynolds Heart Disease
            Prevention grants for use in cDNA microarray experiments. Sequence
            Quality: Sequence ends were trimmed based on percentage of ambigu
            us base calls or 'N's in windowed segments. Sequencing: First-pass
            sequencing; ABI Prism 377 sequencer and analysis software.
            Seq primer: M13/pUC Reverse.
FEATURES              Location/Qualifiers
     source           1. .93
                      /organism="Mus musculus"
                      /db_xref="taxon:10090"
                      /clone="UTSW_SM41H9"
                      /clone_lib="UTSW Adult Mouse Skeletal Muscle Library"
                      /sex="Pooled"
                      /tissue_type="Diaphragm/Hind limb muscles"
                      /cell_type="Skeletal muscle"
                      /dev_stage="2 months"
                      /lab_host="DH5a"
                      /note="Vector: pAMP10 (Gibco); Cloned unidirectionally.
```

```
                Primer: Oligo dT. RNA Isolation: cytoplasmic  RNA preps
                (Manniatis); Cloning Technique: CUA Cloning (CloneAmp,
                Life  Technologies); Average insert size: 1.8 Kb;
                Insertion site:  TACGTCCACTGAATTCTGAGTG--->. Other
                information regarding entire library  may be found at
                http://pga.swmed.edu/Data/Libraries/microarray_cdna_librar
                ies.htm."
BASE COUNT       27 a     22 c      9 g     31 t      4 others
ORIGIN


  Query Match             0.6%;  Score 26;  DB 10;  Length 93;
  Best Local Similarity   57.1%;  Pred. No. 5.1e+04;
  Matches   44;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy   3168 aaatttcagcactcgatataagttttatttctatgcacaaacatcagcaggatcaggaag 3227
          |||  |    ||| ||| || |||| |    |||| | ||| |       || ||||||
Db     77 AAAGCTGGTTACTNGATGTAGGTTTAANAGTTATGTAAAAAAAAANAGCTGACCAGGAAT 18

Qy   3228 tcaaattacagaggaag 3244
          ||  |||  ||  | ||
Db     17 TCGGATTCTAGTAGTAG 1


RESULT   9
BE324593
LOCUS       BE324593                 95 bp    mRNA    linear   EST 14-JUL-2000
DEFINITION  NF014E05PL1F1036 Phosphate starved leaf Medicago truncatula cDNA
            clone NF014E05PL 5', mRNA sequence.
ACCESSION   BE324593
VERSION     BE324593.1  GI:9198370
KEYWORDS    EST.
SOURCE      barrel medic.
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 95)
  AUTHORS   Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores
            ,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula phosphate-starved leaf library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Harrison MJ
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7325
            Fax: 580 221 7380
            Email: mjharrison@noble.org
            Medicago Genome Initiative accession: MGI:S:24060
            Insert Length: 95   Std Error: 0.00
            Plate: 014  row: E  column: 05
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .95
                     /organism="Medicago truncatula"
                     /db_xref="taxon:3880"
                     /clone="NF014E05PL"
                     /clone_lib="Phosphate starved leaf"
                     /tissue_type="leaf"
                     /dev_stage="trifoliate"
                     /note="Vector: Lambda Zap; At the trifoliate stage, M.
                     truncatula plants were transplanted to phosphate-free sand
                     and grown for a further 30 days.  During this 30 day
                     period, the plants were fertilized twice weekly with 1/2
                     Hoaglands solution containing only 20uM potassium
                     phosphate.  RNA was prepared from above ground tissues."
BASE COUNT       27 a     24 c     12 g     32 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 10;  Length 95;
  Best Local Similarity   57.3%;  Pred. No. 5.2e+04;
  Matches   47;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    165 gcccctgattctcttcctgtgccagatgattagtgcactggaagtacctcttgatccaaa 224
          || || |||| ||||| | |||   |   |||||    || |  |  | || |  | |||
Db      1 GCACCAGATTTTCTTCATTTGCTCAACCCTTAGTTGCATGAATATTTCCCTAGCACAAAA 60

Qy    225 acttcttgaagacttggtacag 246
          || | |||| ||  |  ||
Db     61 CTCTCCTCAAGATTTTCTTGAG 82


RESULT  10
W03279/c
LOCUS       W03279                   96 bp    mRNA    linear   EST 18-APR-1996
DEFINITION  za36c11.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:294644 5' similar to gb:M27492 INTERLEUKIN-1 RECEPTOR, TYPE I
            PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION   W03279
VERSION     W03279.1  GI:1272741
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 96)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Trace considered overall poor quality
            Seq primer: ETPrimer
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .96
                     /organism="Homo sapiens"
                     /db_xref="GDB:1239565"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:294644"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       26 a     11 c     14 g     29 t     16 others
ORIGIN


  Query Match             0.6%;  Score 26;  DB 10;  Length 96;
  Best Local Similarity   49.4%;  Pred. No. 5.2e+04;
  Matches   44;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   3123 tttgaaaattcctgccaacaagacacggtggactttaaaaaatttaaatttcagcactcg 3182
          | ||  || ||||||  ||||  | | || | ||  |    |   |   |   |   |
```

```
Db      94 TGTGGNAAATCCTGCAAACANNANAAGGAGTACCCTNNTCACNNTGCTTAATATNTCGNA 35

Qy    3183 atataagttttatttctatgcacaaacat 3211
           || | |   |   || |||  |||  |||
Db      34 ATNTNANANTAGATTTTATATACATNCAT 6


RESULT 11
AW845547/c
LOCUS       AW845547                100 bp    mRNA    linear   EST 19-MAY-2000
DEFINITION  MR1-CT0058-291199-003-d02 CT0058 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW845547
VERSION     AW845547.1  GI:7941064
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR1-CT0058-291
            199-003-d02&t3=1999-11-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 30
            High quality sequence stop: 99.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CT0058"
                     /dev_stage="Adult"
                     /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT       17 a     34 c     11 g     38 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 9;  Length 100;
  Best Local Similarity   57.3%;  Pred. No. 5.3e+04;
  Matches   47;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     775 agatttaatcatactcaaaccatacagcagaagcaacctatttctgtgaaggtgatttca 834
           |||| |||| |    |||| |   ||| ||| |  | | |     |||||||  | |  |
Db      82 AGATGTAATAAGGAACAAAGCTAGCAGGAGAGGAGAGCCAGAGAAGTGAAGGGAAGTCTA 23

Qy     835 gtggatgaattgaatgacacta 856
           || |  || || ||  ||||
Db      22 CAGGTTACATAGATTGTGACTA 1


RESULT 12
D45290/c
LOCUS       D45290                  100 bp    mRNA    linear   EST 30-DEC-1995
DEFINITION  HUMHG2360 Human cerebral cortex Homo sapiens cDNA, mRNA sequence.
ACCESSION   D45290
VERSION     D45290.1  GI:1136686
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100)
  AUTHORS   Takahashi,N., Hashida,H., Zhao,N., Misumi,Y. and Sakaki,Y.
  TITLE     High-density cDNA filter analysis of the expression profiles of the
            genes preferentially expressed in human brain
  JOURNAL   Gene 164, 219-227 (1995)
  MEDLINE   96069586
COMMENT     Contact: Nobuaki Takahashi
            Institute of Medical Science
            University of Tokyo
            Shirokanedi 4-6-1, Minato-ku, Tokyo, Japan 108
            Tel: 03-5449-5625
            Fax: 03-5449-5445.
FEATURES             Location/Qualifiers
     source          1. .100
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human cerebral cortex"
                     /note="Adult male cerebral cortex tissue."
BASE COUNT       35 a     19 c     18 g     28 t
ORIGIN


  Query Match             0.6%;  Score 26;  DB 10;  Length 100;
  Best Local Similarity   59.5%;  Pred. No. 5.3e+04;
  Matches   44;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     853 actatagctgctaatttgagtgacactgagttttatggtgctaaatcaagtagagagagg 912
           || | | | |||  | | |||      |  ||  || | |   |||||| | || | |||
Db      79 ACCAAAACAGCTTTTATTAGTATGGATTTGTAATAAGTTTTGAAATCAGGAAGTGTGAGT 20

Qy     913 ccaccaacattttt 926
           ||||||| || ||
Db      19 TCACCAACTTTGTT 6


RESULT 13
AZ826437
LOCUS       AZ826437                71 bp    DNA     linear   GSS 20-FEB-2001
DEFINITION  2M0102K08F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC2M0102K08 F, DNA sequence.
ACCESSION   AZ826437
VERSION     AZ826437.1  GI:12996345
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
```

```
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0102  row: K  column: 08
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 71.
FEATURES              Location/Qualifiers
     source           1. .71
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC2M0102K08"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       17 a     10 c     17 g     27 t
ORIGIN


  Query Match             0.6%;  Score 25.8;  DB 12;  Length 71;
  Best Local Similarity   60.9%;  Pred. No. 5e+04;
  Matches   42;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy   2386 tcagagcctgataatttggagattacgtggaagcccttgaatggtttcgaatctaatggg 2445
           | || || || | || || ||  ||| ||| |    || |||  ||| ||||||  || |
Db      3 TTAGTGCATGGTGATCTGTAGGATACATGGGACAATTTCAATATTTTTGAATCTGTTGAG 62

Qy   2446 ccaggcctt 2454
           |  |  ||
Db     63 GCCTGATTT 71


RESULT  14
BE289162
LOCUS       BE289162                 85 bp    mRNA    linear   EST 26-OCT-2000
DEFINITION  601092758F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3487430 5',
            mRNA sequence.
ACCESSION   BE289162
VERSION     BE289162.1  GI:9169349
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 85)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8525  row: c  column: 15
            High quality sequence stop: 85.
FEATURES              Location/Qualifiers
     source           1. .85
                      /organism="Mus musculus"
                      /strain="C57/B6"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:3487430"
                      /clone_lib="NCI_CGAP_Mam5"
                      /tissue_type="tumor, gross tissue"
                      /dev_stage="7 months"
                      /lab_host="DH10B"
                      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                      Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                      Library constructed by Life Technologies. Investigators
                      providing samples: Lothar Hennighausen/Robin Humphreys,
                      NIH"
BASE COUNT       35 a     10 c     22 g     18 t
ORIGIN


  Query Match             0.6%;  Score 25.8;  DB 10;  Length 85;
  Best Local Similarity   63.9%;  Pred. No. 5.5e+04;
  Matches   39;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy   3410 aatatggtgtagcaggcagcaaagaagaatggagaaaagaaattgtaaatggttctcgga 3469
           || | || | || | || ||| || |  |||  |||||||||   ||| |   | | |||
Db      5 AAAAGGGCGAAGGAAGCTGCAGAGCAAGATGTTGAAAAGAAAAAATAATTCTGTGTTGGA 64

Qy   3470 g 3470
           |
Db     65 G 65


RESULT  15
FR0022001
LOCUS       FR0022001                99 bp    DNA     linear   GSS 10-DEC-1997
DEFINITION  F.rubripes GSS sequence, clone 070J16aE2, genomic survey sequence.
ACCESSION   AL014872
VERSION     AL014872.1  GI:2681240
KEYWORDS    GSS; genome survey sequence.
SOURCE      Takifugu rubripes.
  ORGANISM  Takifugu rubripes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Takifugu.
REFERENCE   1  (bases 1 to 99)
  AUTHORS   Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y.,
            Williams,G. and Brenner,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
            Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hgmp.mrc.ac.uk
COMMENT     Vector: pBluescript II KS
            V_type: phagemid
            PRIMER: KS
            DESCR:
            One pass dye-terminator sequencing of cosmid cloned genomic
            sequence.
FEATURES              Location/Qualifiers
     source           1. .99
                      /organism="Takifugu rubripes"
                      /db_xref="taxon:31033"
                      /clone_lib="cosmid 070J16"
                      /clone="070J16aE2"
BASE COUNT       24 a     26 c     23 g     24 t      2 others
ORIGIN
```

Query Match 0.6%; Score 25.8; DB 12; Length 99;
Best Local Similarity 57.0%; Pred. No. 6e+04;
Matches 45; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy   1905 caaggatcatctagtggtagctgatgtcagtgacgatgacagcgggacctacacgtgtgt 1964
          ||| |     || ||  ||     |||||     || ||    |||  |||||| ||
Db     16 CAATGGAACCCTGGTCCTACGATCTGTCAAANCAGAGGATTCTGGGTACTACACATGCNC 75

Qy   1965 ggccaacaccactctggac 1983
          |||| || |||||||| |
Db     76 TGCCACCAACACTCTGGGC 94
```

Search completed: August 13, 2002, 21:18:23
Job time: 6465 sec